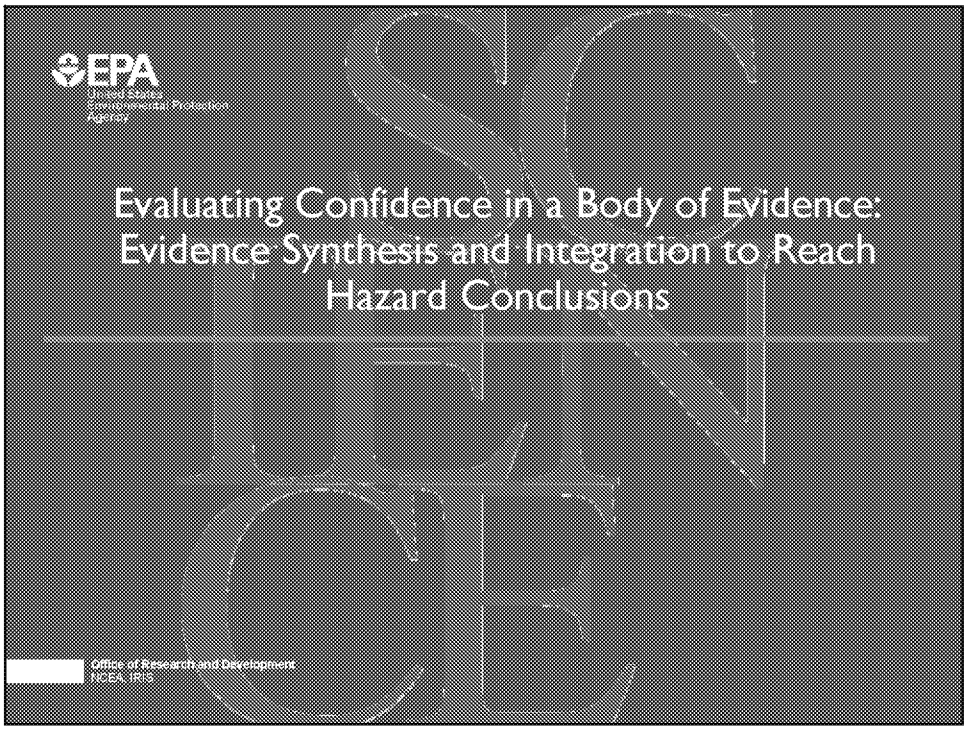
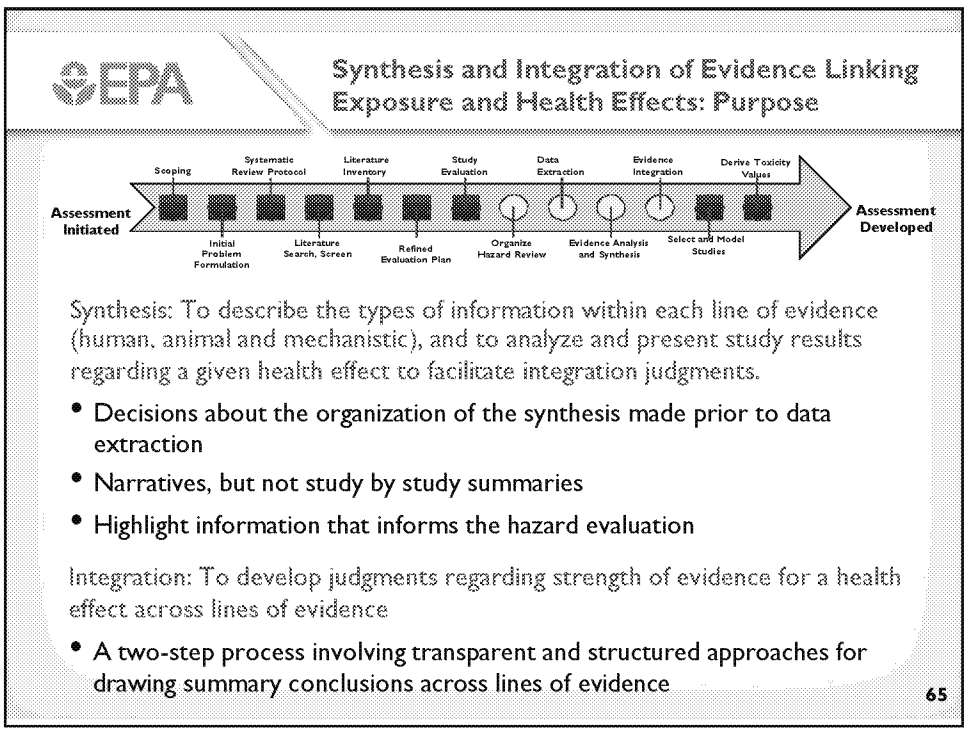




Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

NAS 2014: Relevant Comments and Recommendations

The NAS 2014 report discusses the complexities with organizing analyses around mechanism, noting that, "The history of science is replete with solid causal conclusions in advance of solid mechanistic understanding." (NRC, 2014, p. 90).

- The current approach focuses first on the available human and animal studies on health effects, incorporating mechanistic information at various stages of assessment development to clarify identified gaps in understanding (e.g., human relevance of animal-model data).

"The risk-of-bias assessment of individual studies should be carried forward and incorporated into the evaluation of evidence among data streams." (NAS 2014 Recommendation, Box 8-1)

- The results of the evaluation of individual studies is a critical component of the current evidence synthesis processes and integration frameworks.

66

NAS 2014: Relevant High Priority (Box 8-1) Recommendations

"**EPA should continue to improve its evidence-integration process incrementally and enhance the transparency of its process.** It should either maintain its current guided-expert-judgment process but make its application more transparent or adopt a structured (or GRADE-like) process...the committee does not offer a preference but suggests that EPA consider which approach best fits..."

"**EPA should expand its ability to perform quantitative modeling of evidence integration.**"

- The current approach continues to use a guided expert judgment process, but structured sets of categorical criteria for decision-making within that process are more explicitly defined.
- The current frameworks, and documentation of decisions within these frameworks, enhance transparency, reproducibility, and comparability across health effects and assessments; these approaches are evolving within NCEA and across the field.
- Current research activities include quantitative methods to integrate evidence across streams (e.g., Bayesian approaches; see Session 4)

67

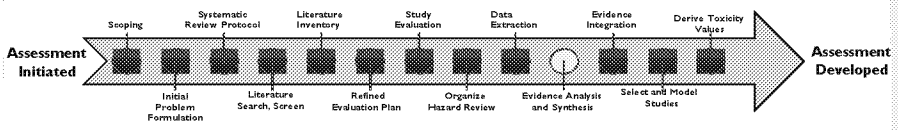
Synthesizing Evidence on Health Effects – Organization and Structure

Some questions about the evidence

- What outcomes are relevant to each health hazard domain and at what level (e.g., health effect or subgroupings) should synthesis occur?
- What populations were studied (e.g., general population, occupations, life stages, species, etc.) and do responses vary?
- Can study results be described across varying exposure patterns, levels, duration or intensity?
- Are there differences in the confidence in study results for different outcomes, populations, or exposure?
- Does toxicokinetic information explain differences in responses across route of exposure, other aspects of exposure, species, or life stages?
- How might dose response relationships be presented (specific study results or across study results)?

68

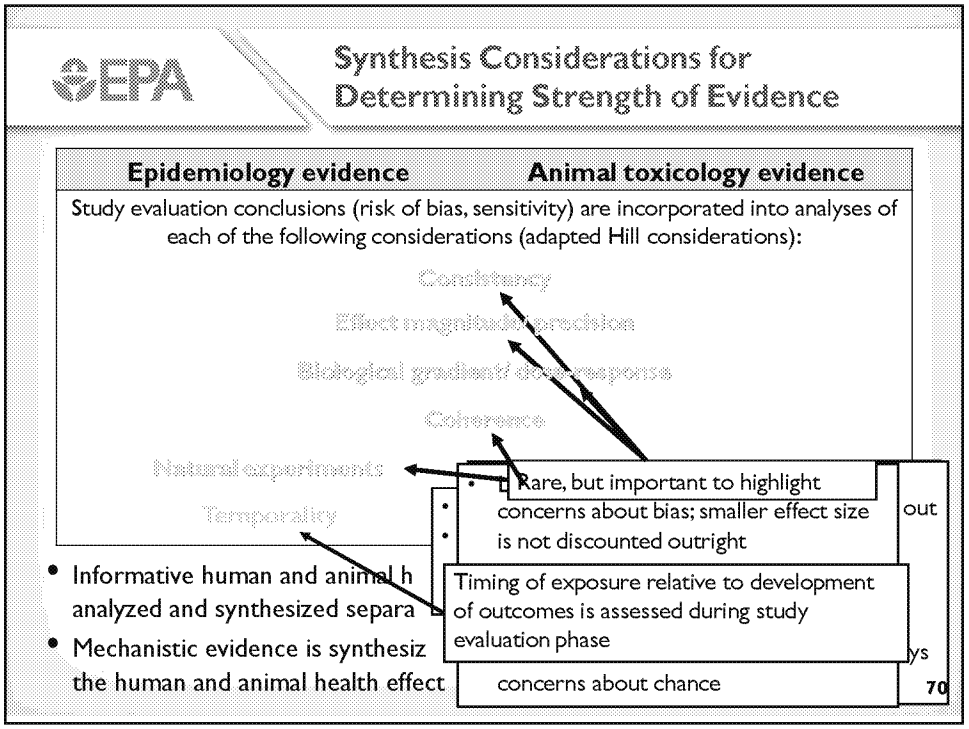
Scientific Judgment in Analysis and Synthesis of Evidence

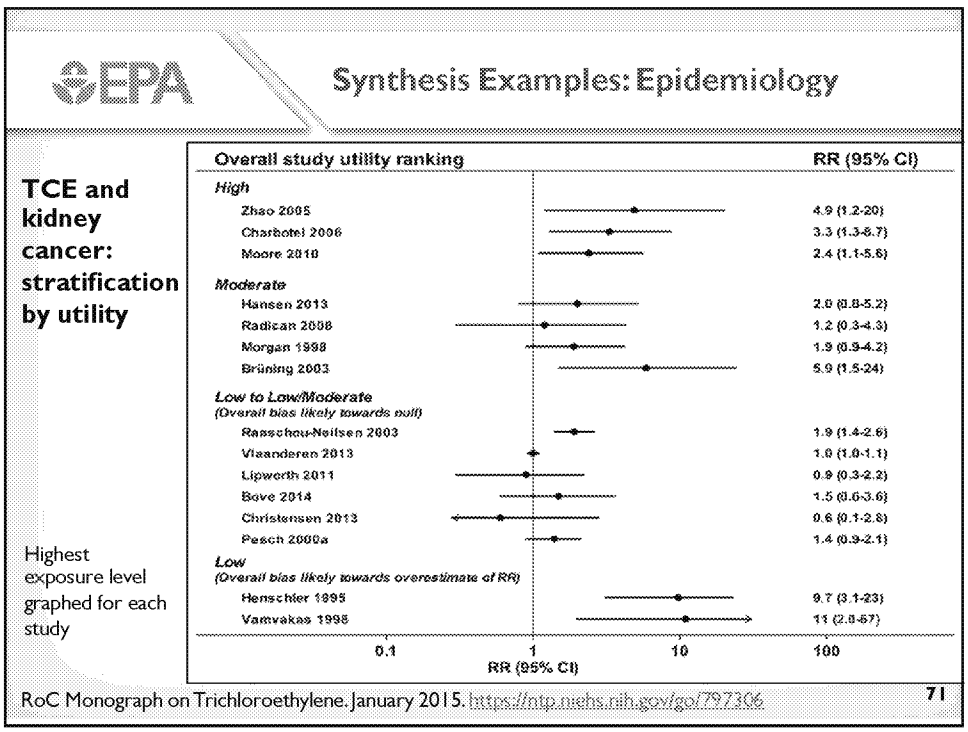


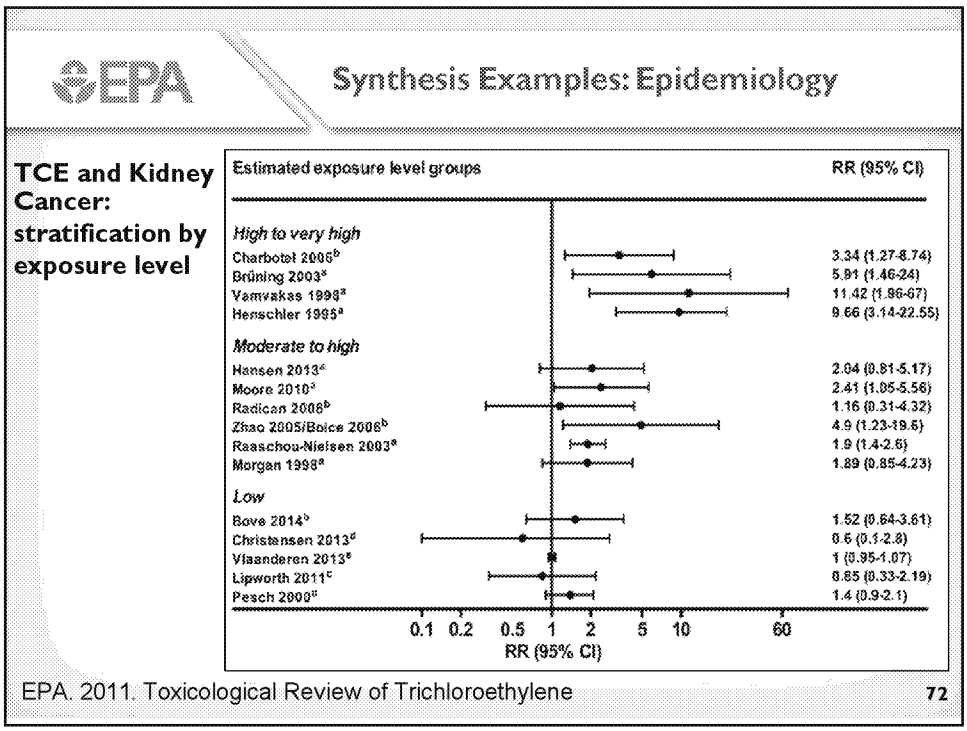
- Synthesis of evidence is more than counting the number of "positive" and "negative" studies
- Must systematically consider the influence of bias and sensitivity when describing study results and synthesizing evidence
- Synthesis should primarily be based on studies of medium and high confidence (when available)
- Analysis should try to draw conclusions about the strength of evidence from findings across collections of studies

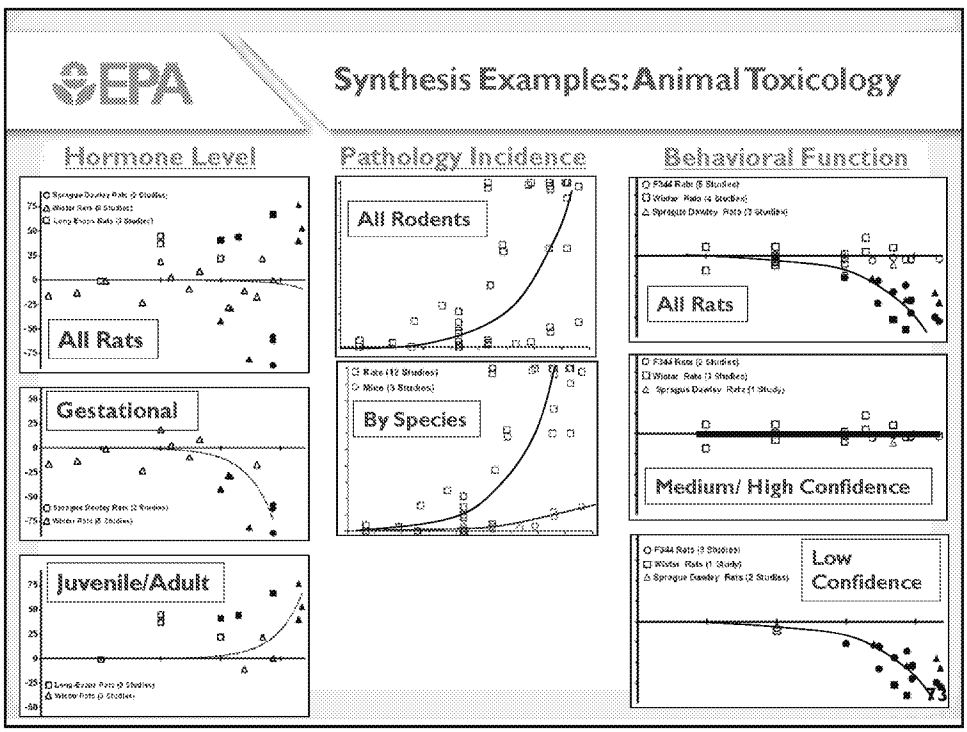
69

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation









Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

EPA

Mechanistic Evidence

"Mechanistic data represent a wide variety of studies not intended to identify an adverse outcome." (NRC, 2014)

- When evaluating mechanistic evidence, the scope is larger than "*in vitro*" data
- Mechanistic inventories collected at earlier stages may include:
  - *In vivo* (cellular, biochemical, molecular)
  - *In vitro* or *ex vivo* (human or animal tissues or cells)
  - Non-animal or non-mammalian alternative animal models
  - Big data ('omics or high-throughput assays)
  - "Intervention" studies (pharmacologic, environmental, genetic)

"…there might be hundreds of *in vitro* and other mechanistic studies of a given chemical…" (NRC, 2014)

"For a given chemical, multiple mechanisms might be involved in a given end point, and it might not be evident how different mechanisms interact in different species to cause the adverse outcome." (NRC, 2014)

74

EPA

Systematic review of mechanistic information requires a different approach

"When human data are nonexistent, are mixed, or consistently show no association and an animal study finds a positive association, the importance of mechanistic data is increased..." (NRC, 2014)

To narrow the scope of the analyses of mechanistic information, IRIS applies an iterative approach to identifying key mechanistic questions at various stages of the systematic review

- *Problem formulation* identifies predefined analyses (e.g., when a mutagenic MOA is indicated)
- *Literature inventory* allows identification of studies on an organ system that human and animal studies meeting the PECO criteria have not examined
- *Human and animal evidence syntheses* may flag impactful qualitative and quantitative analyses

75

Human and animal evidence syntheses may flag impactful mechanistic analyses

- Identify precursor events for apical toxicity endpoints
- Inform susceptibility (species, strain, or sex differences; at-risk populations or lifestages)
- Inform human relevance of animal data (note: the level of analysis will vary depending on the impact of the animal evidence)
- Provide biological plausibility (i.e., to human or animal health effect data when evidence is weak or critical uncertainties are identified)
- Establish mechanistic relationships (or lack thereof) across sets of potentially related endpoints/outcomes to inform the consideration of coherence during evidence integration
- Aid extrapolation (high-to-low dose; short-to-long duration; route-to-route)
- Improve dose-response modeling and quantification of uncertainties

76

Mechanistic Analysis Focused on Specific Questions

Examples of when these analyses have been triggered in recent IRIS Assessments:

- Benzo[a]pyrene (2017): The descriptor "carcinogenic to humans" was supported by strong mechanistic evidence that established the biological plausibility of the animal findings occurring in humans, despite lack of human exposure data
  - Key precursors (BPDE-DNA adducts) were identified in humans exposed to PAH mixtures that are specific to B[a]P, form mutational spectra unique to B[a]P, and are associated with cancer in humans
- Dichloromethane (2011): The cancer risk estimate was specifically derived for a susceptible subpopulation (GSTT1+/+) identified by the mechanistic evaluation
  - Differing results *in vivo* were explainable by species and tissue differences in the availability of GST
  - PBPK modeling addressed the variability in this population
- Documentation and transparency is key for future mechanistic analyses

77

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

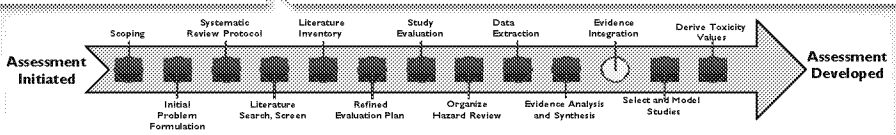
Focused mechanistic evaluations

"Several criteria should be considered in assessing in vitro toxicology studies for risk of bias and toxicologic relevance. Relevance should be determined in several domains, including cell systems used, exposure concentrations, metabolic capacity, and the relationship between a measured in vitro response and a clinically relevant outcome measure. **Few tools are available for assessing risk of bias in in vitro studies. Because of the nascent status of this field, the committee can provide only provisional recommendations for EPA to consider...**EPA should carry out, support, or encourage research on the development and evaluation of empirically based instruments for assessing bias in…mechanistic studies." (NRC, 2014)

- Prioritize studies of relevant endpoints and associated assays by toxicologic relevance (e.g., model systems; dose range; sensitivity and specificity of assay)
- Conduct individual study evaluations on the most impactful studies
- EPA is exploring the use of existing tools, including adaptations of IRIS study evaluation tools
- Organizational frameworks (e.g., EPA's MOA framework using modified Hill considerations; visual AOP-like constructs) are useful for organizing and documenting these analyses transparently to convey conclusions for evidence integration

78

Moving from Synthesis to Integration



Outputs of Evidence Synthesis

Results of Human Health Effect Study Synthesis

Results of Animal Health Effect Study Synthesis

Results of Synthesis of Mechanistic Evidence Informing the Human and Animal Syntheses

Evidence Integration

Transparent and Structured Processes for Drawing Summary Conclusions Across Lines of Evidence

79

Evidence Integration Involves a Sequential, Two-Step Process

- **Evidence synthesis interpretations for each consideration relating to causality are combined across lines of evidence using transparent, structured frameworks**

Step 1: "Within-Stream" Integration	Step 2: "Across-Stream" Integration
Judge the Strength of the Evidence from the: • **Human Evidence Stream** • **Animal Evidence Stream**	Draw Overall Evidence Integration Conclusions based on: • Combined **Human** and **Animal Evidence Streams**
Human health effect study synthesis conclusions for each consideration are integrated in light of mechanistic evidence in exposed humans or human cells (or other human models)	The judgments regarding the strength of the human and animal evidence streams are integrated in light of evidence on the human relevance of the findings in animals, susceptibility, and the coherence of the findings across evidence streams.
Characterize the Strength of the Evidence for an Effect in Animals (**Animal Evidence Stream Judgment**)	
Animal health effect study synthesis conclusions for each consideration are integrated in light of mechanistic evidence in exposed animals or animal cells (or other relevant models)	

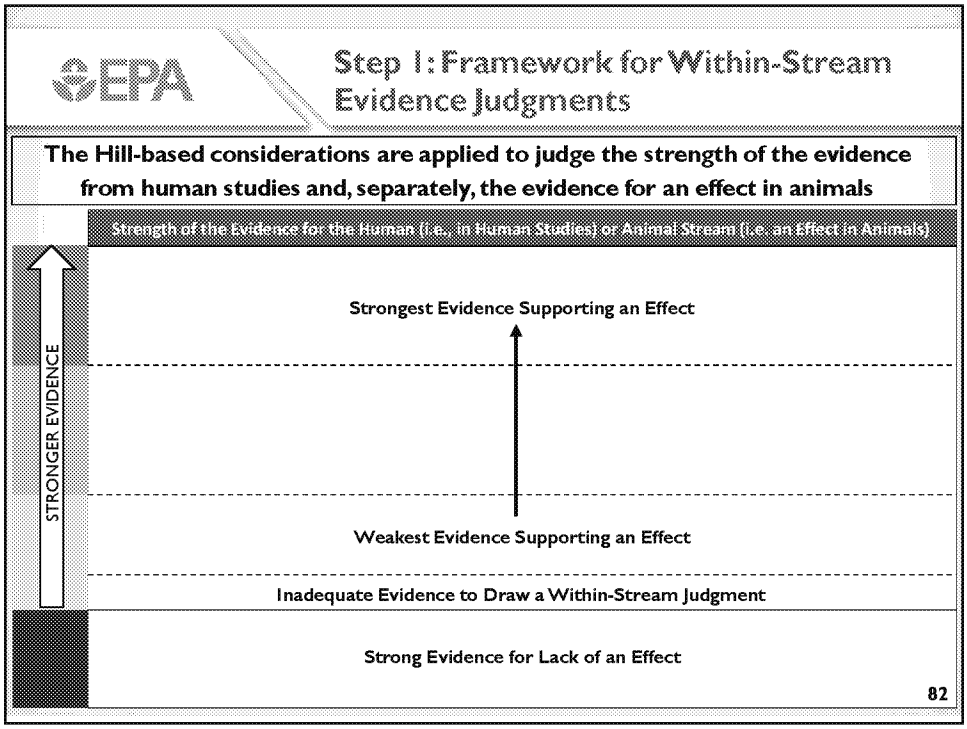
80

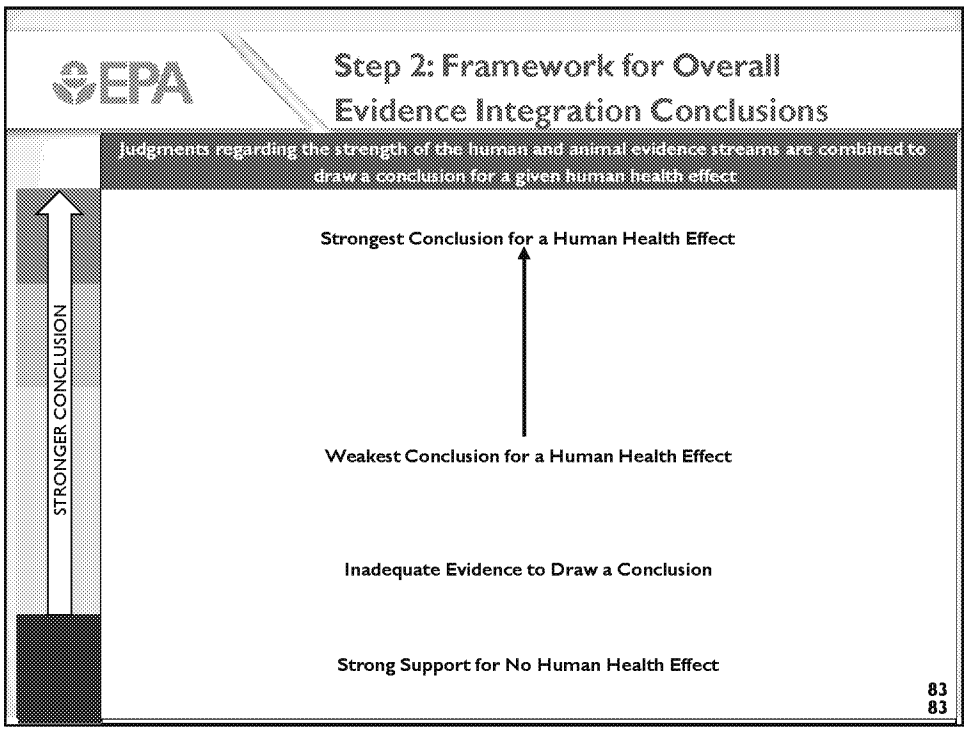
Within-Stream (Human; Animal Stream) Evidence Judgment Considerations

	Human Evidence Stream	Animal Evidence Stream
Individual Studies	• *High* or *medium* confidence studies provide stronger evidence within evaluations of each Hill consideration • Interpreting results considers biological as well as statistical significance, and findings across studies	
Consistency	• Different studies or populations increase strength	• Different studies, species, or labs increase strength
Dose-response	• Simple or complex (nonlinear) relationships provide stronger evidence • Dose-dependence that is expected, but missing, can weaken evidence (after considering the findings in the context of other available studies and biological understanding)	
Magnitude, Precision	• Large or severe effects can increase strength; further consider imprecise findings (e.g., across studies) • Small changes don't necessarily reduce evidence strength (consider variability, historical data, and bias)	
Coherence	• Biologically related findings within an organ system, within or across studies, or across populations (e.g., sex) increases evidence strength (considering the temporal- and dose-dependence of the relationship) • An observed lack of expected changes reduces evidence strength	
	• Informed by mechanistic evidence on the biological development of the health effect or toxicokinetic/ dynamic knowledge of the chemical or related chemicals	
Mechanistic Evidence on Biological Plausibility	• Mechanistic evidence in humans or animals of precursors or biomarkers of health effects, or of changes in established biological pathways or a theoretical mode-of-action, can strengthen evidence • Lack of mechanistic understanding does not weaken evidence outright, but it can if well-conducted experiments exist and demonstrate that effects are unlikely	

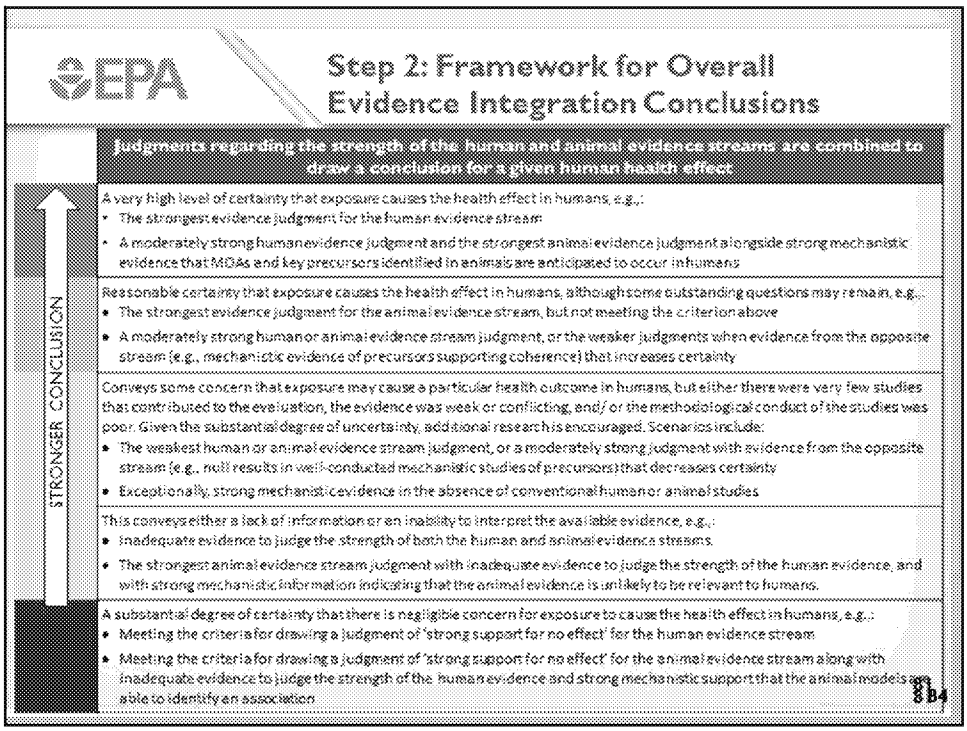
Light blue rows highlight mechanistic inferences; "temporality" and "natural experiments" not shown

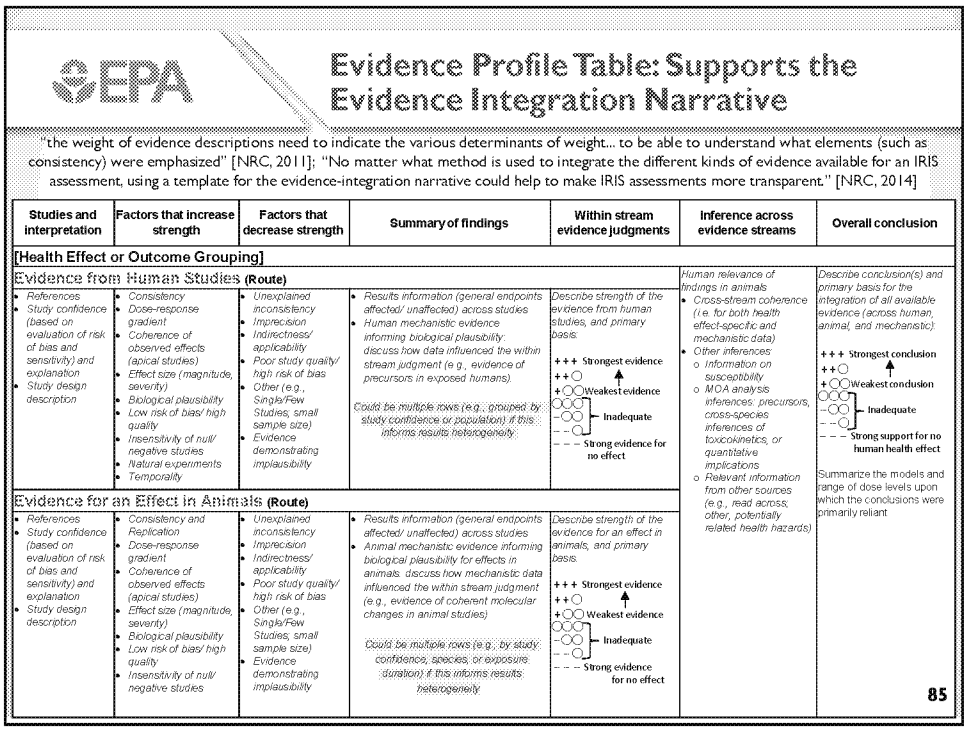
Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation





Appendix C





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Evidence Integration Conclusions

- For Cancer, conclusions on the integrated evidence for each cancer type (or grouping) are evaluated in the context of MOA information to develop an evidence integration narrative that includes a descriptor for carcinogenicity:
  - **carcinogenic** to humans; **likely** to be carcinogenic to humans; **suggestive** evidence of carcinogenic potential; **inadequate** information to assess carcinogenic potential; or **not likely** to be carcinogenic to humans
- For Noncancer Effects, frameworks for evaluating the integrated evidence have been developed to add structure and transparency to the evidence integration narrative(s), which include(s) the relevant exposure context.
  - IRIS has not yet incorporated standardized descriptors for noncancer effects
  - The NAS recommended incremental improvements in this area, including recommendations to "Develop uniform language to describe strength of evidence on noncancer effects" [p. 92, 2014]
  - The specific way in which these conclusions are summarized is currently being tested and discussed within EPA

86

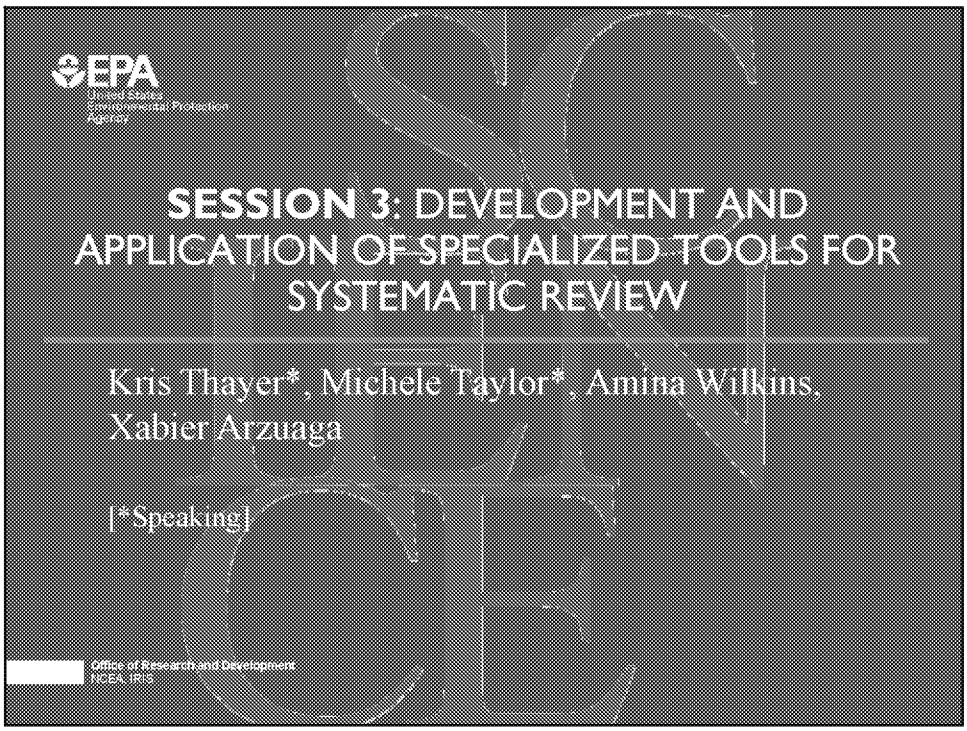
IRIS has Addressed the Major NAS 2014 Recommendations

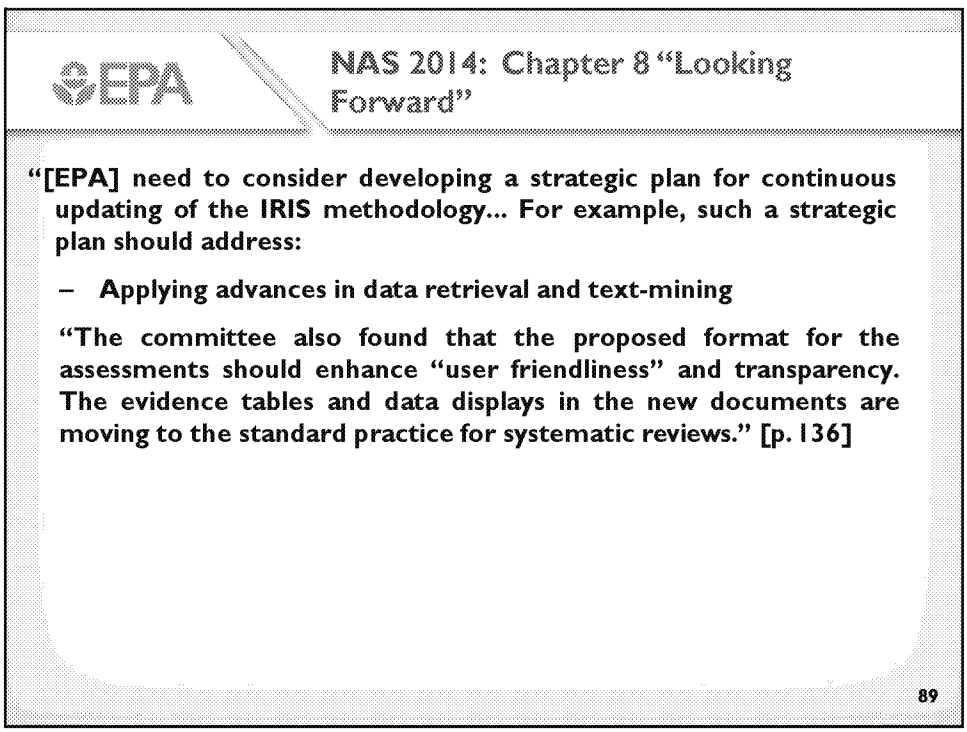
NAS 2014 Topics	IRIS Process Improvements
Evidence Evaluation (Chapter 5)	• Individual studies are evaluated for reporting quality, risk of bias, and sensitivity • Decisions and supporting rationale are clearly documented • Study evaluations impact subsequent assessment decisions
Evidence Integration for Hazard Identification (Chapter 6)	• Structured frameworks provide transparency in expert judgments across human, animal, and mechanistic studies (based on Hill) • Standardized templates documenting key evidence integration decisions have been developed (evidence profile tables)

See Posters and Demonstrations:

- Male reproductive toxicity in studies of phthalates (4 posters on a case study for each of the 3 lines of evidence and the overall evidence integration)
- Combining data within species (poster on meta-analytical approaches)
- PBPK model evaluation for human health assessments (poster)
- Health Assessment Workspace Collaborative (demonstration)

87





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Current Application of Systematic Review Software

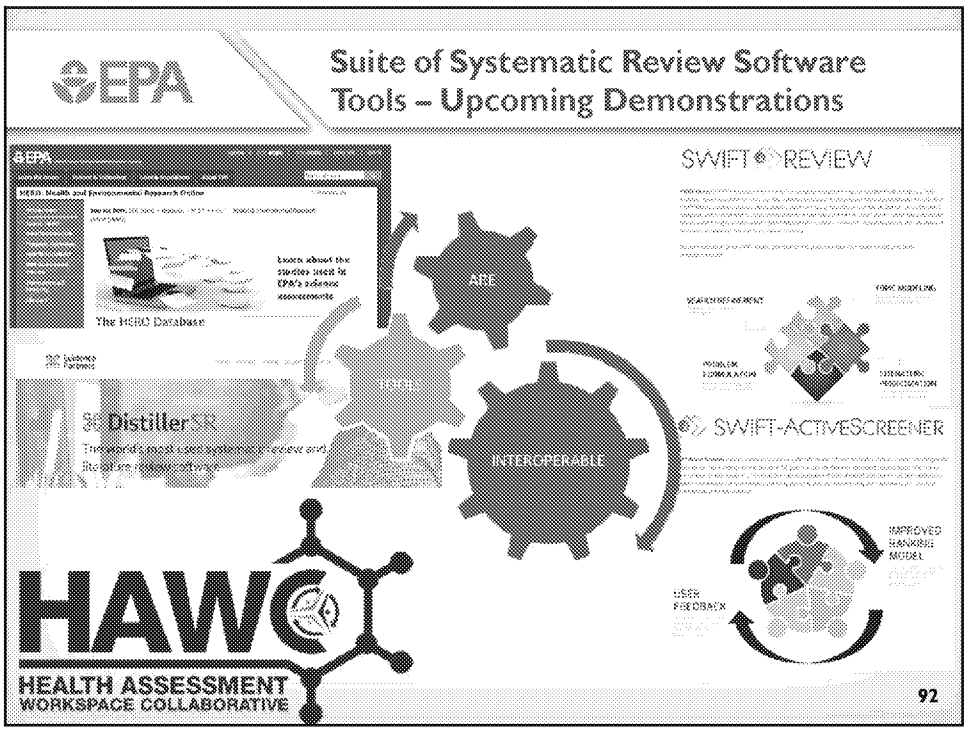
- Specialized software tools make the process more efficient
  - Time and cost savings, improved data management, increased transparency
- NOT all systematic review software tools are intended to automate/semi-automate the process, e.g., HAWC helps manage information content
  - Currently, automation tools are most advanced for evidence identification
- Prefer free tools when possible to help address needs of a potentially large community of users in environmental and biomedical sciences
- Incorporate tools after confirming acceptable performance and interoperability with HERO
  - A toolbox approach, not a "one and only" tool model
- Organized multiple IRIS staff training sessions in 2017 and created a support team ("train the trainers" model)

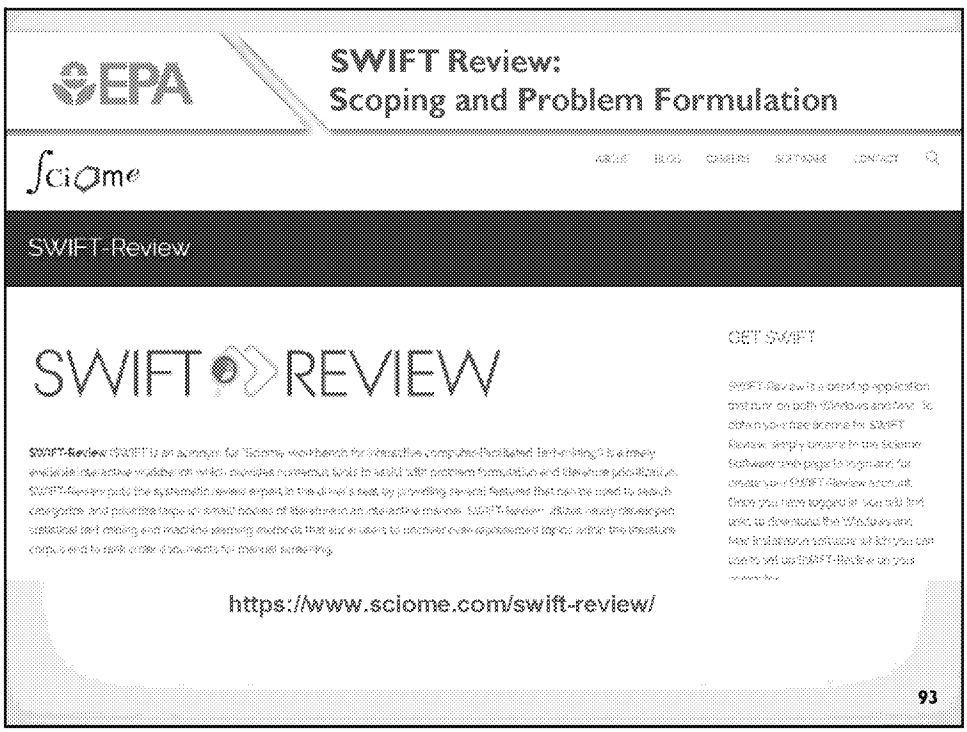
90

Research Activities

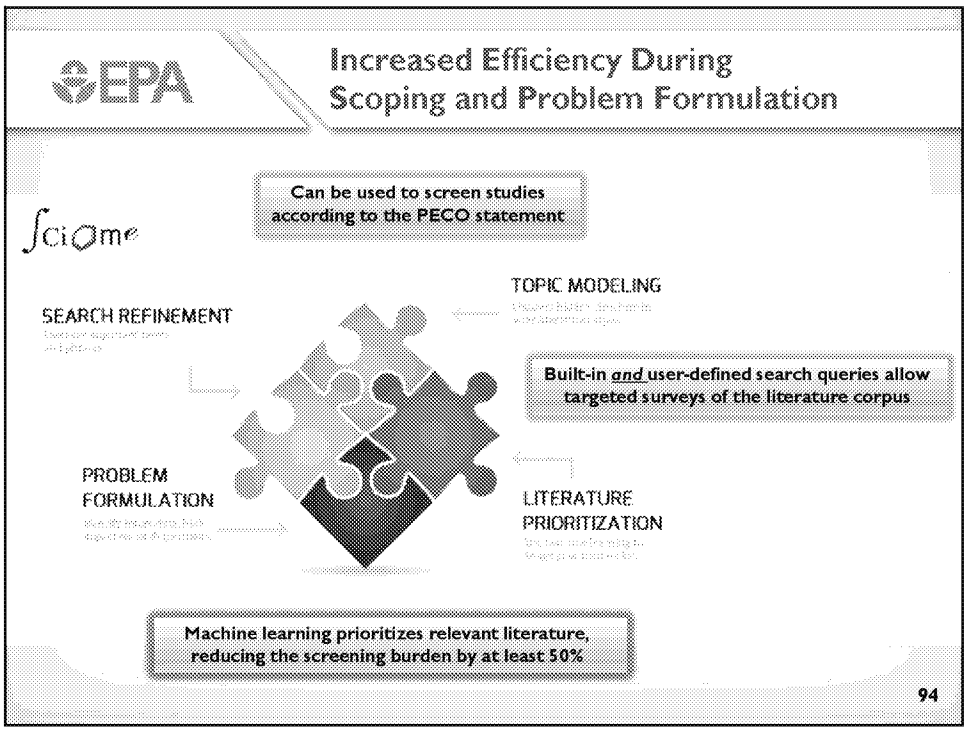
- Developing tools to help automate beyond evidence identification is a long-term research commitment
  - Major hurdle is lack of training/test sets for model development
  - Better performance expected for more structured content (e.g., animal bioassay compared to epidemiological studies)
- Any progress on semi-automation could result in large time and cost savings
- In 2017, NCEA created an interagency agreement with NTP to leverage resources
  - Current activities focus on creating test/training sets and model development for basic content of animal studies (e.g., test chemical, species, dose levels, randomization, etc.).
  - Other parts of EPA can also utilize interagency agreement
- Innovation challenges may be required to identify solutions for capturing complex content, i.e., table content, information spread across multiple sentences and paragraphs

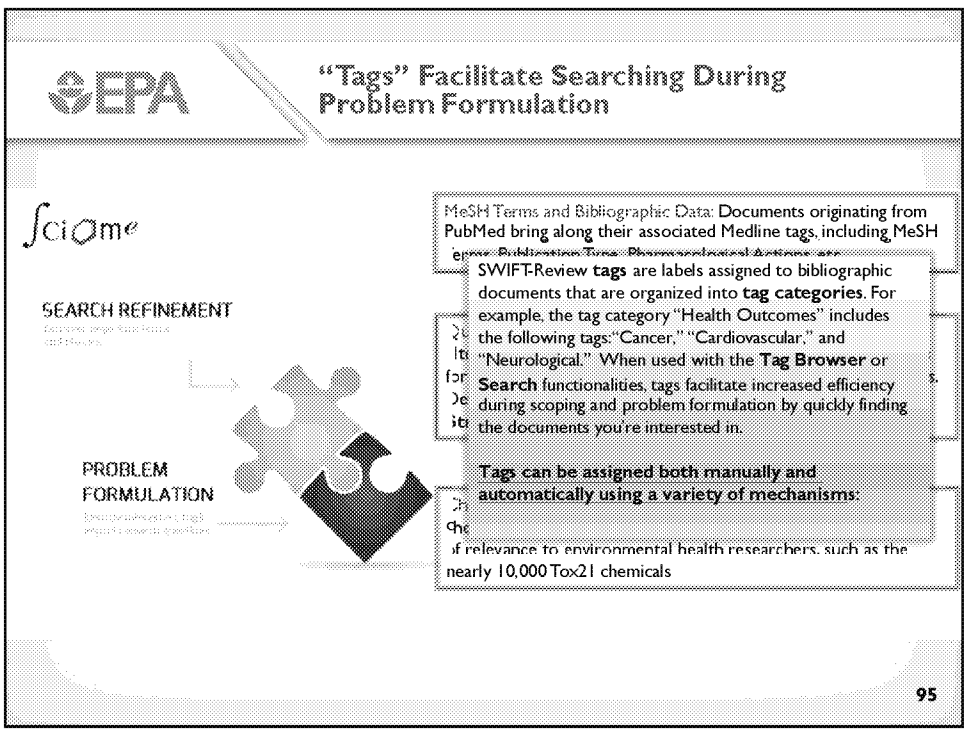
91

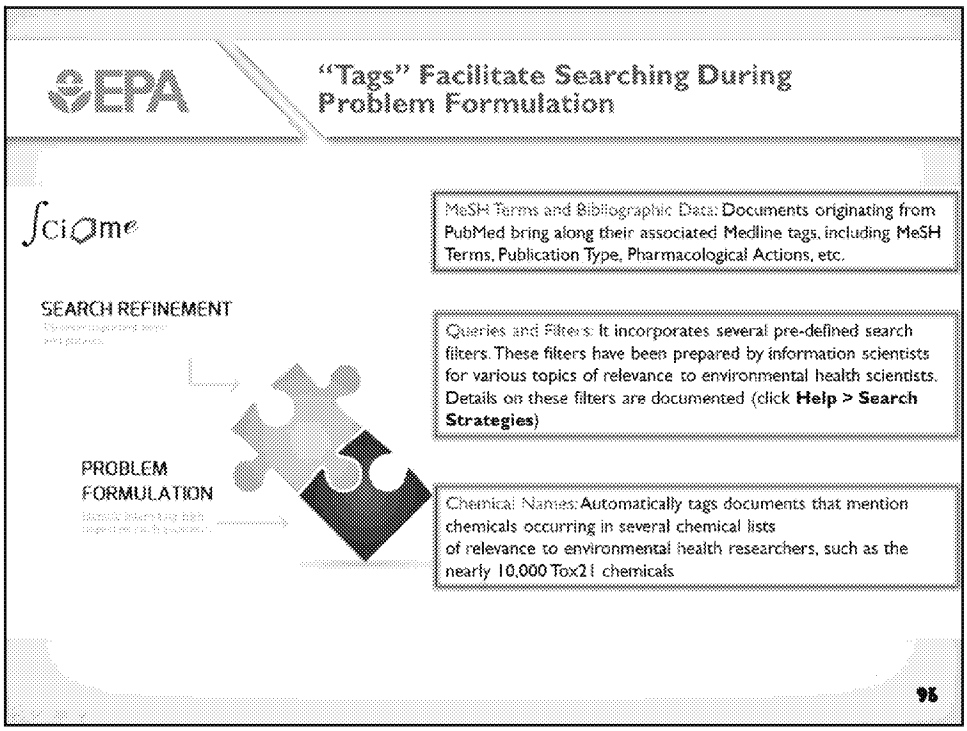


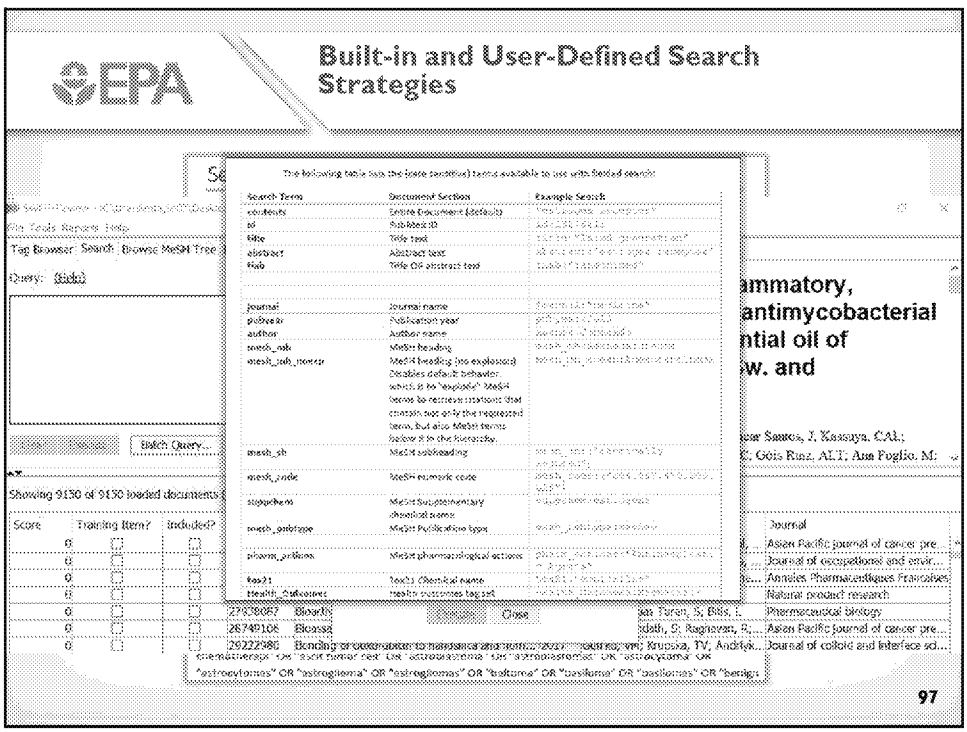


Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

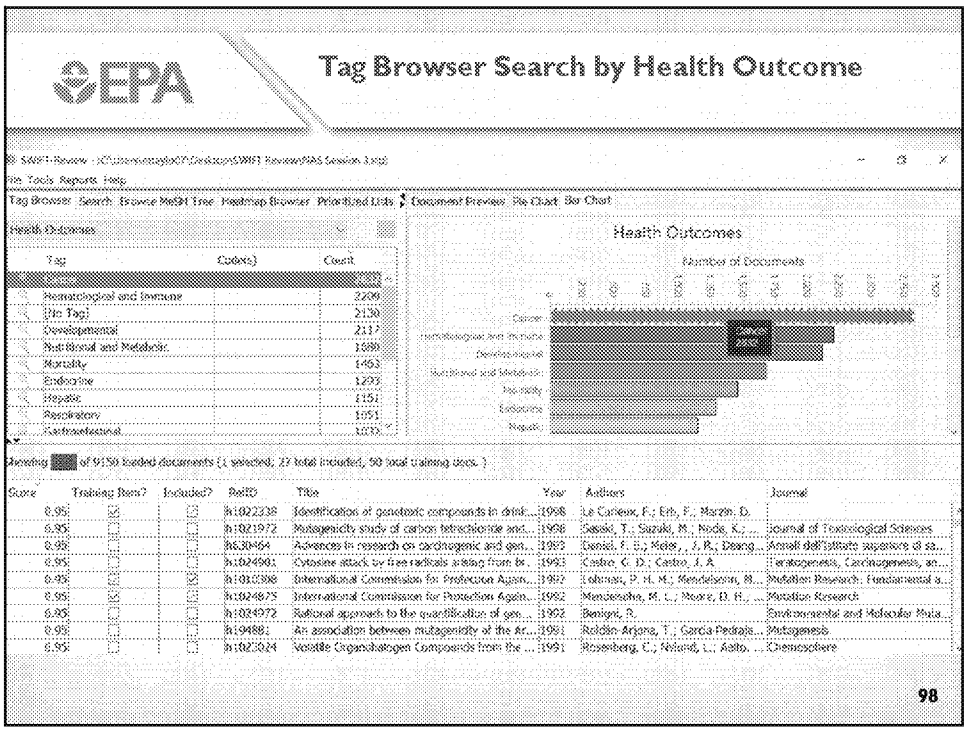


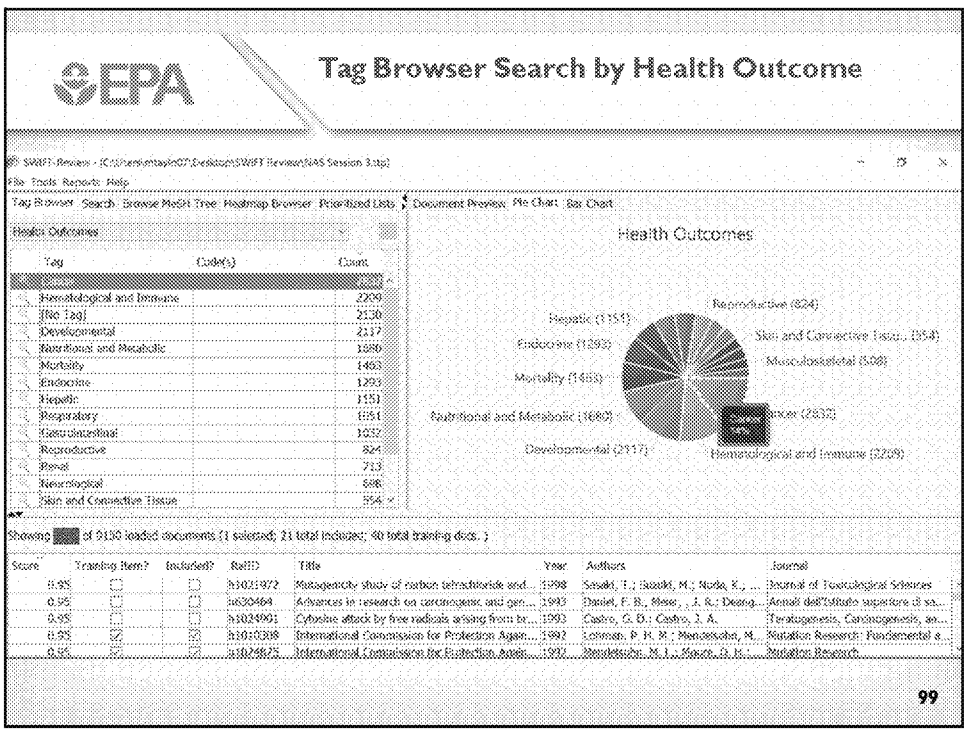




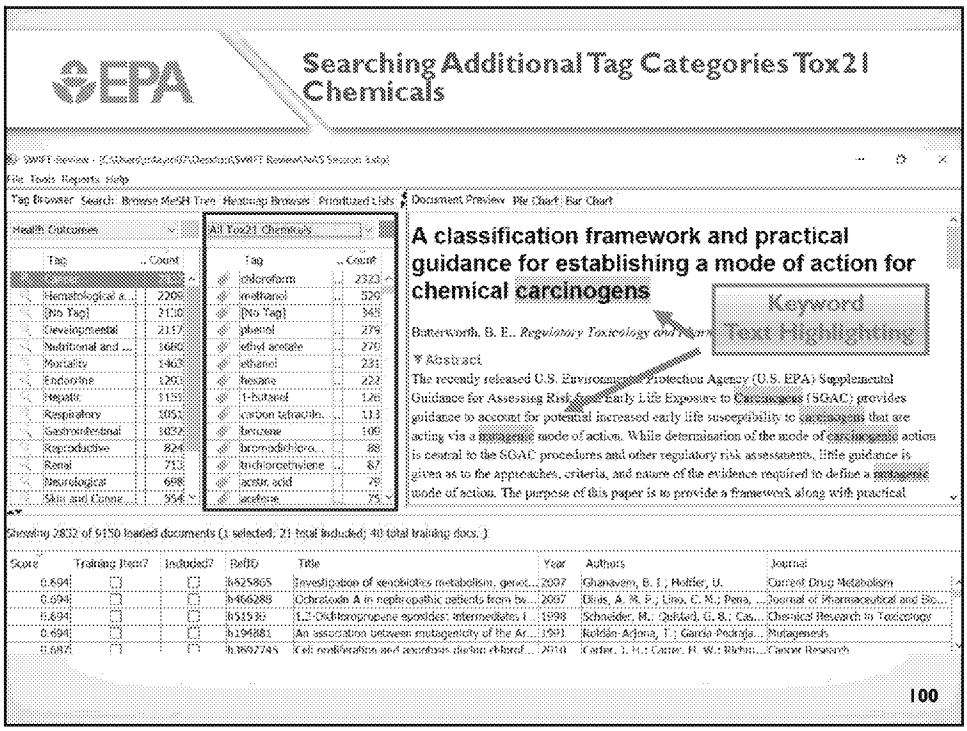


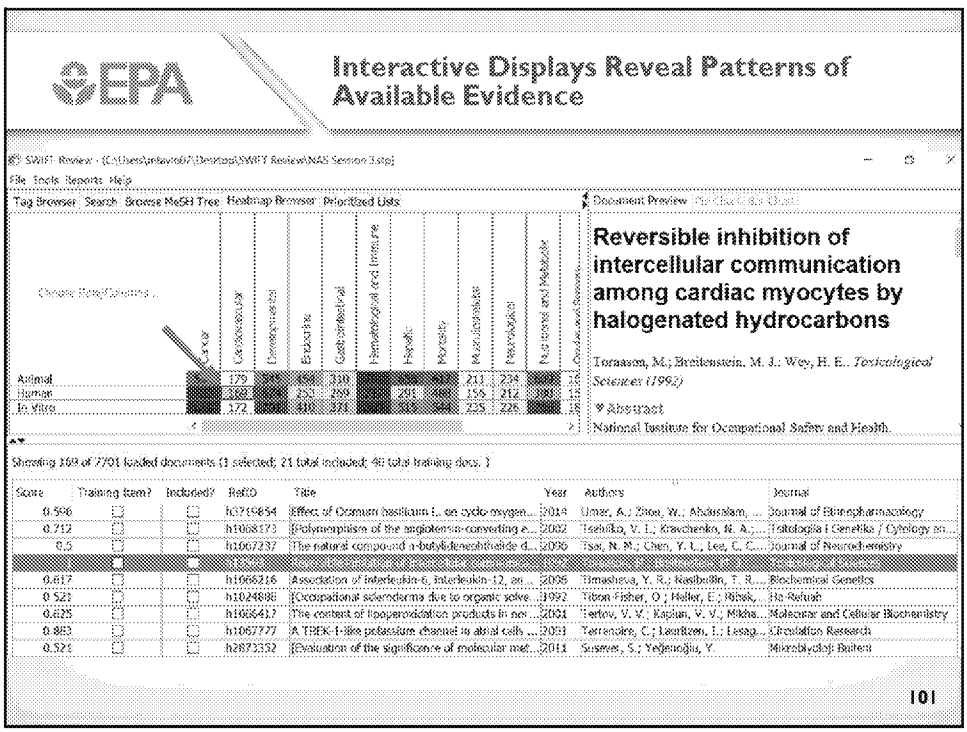
Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation



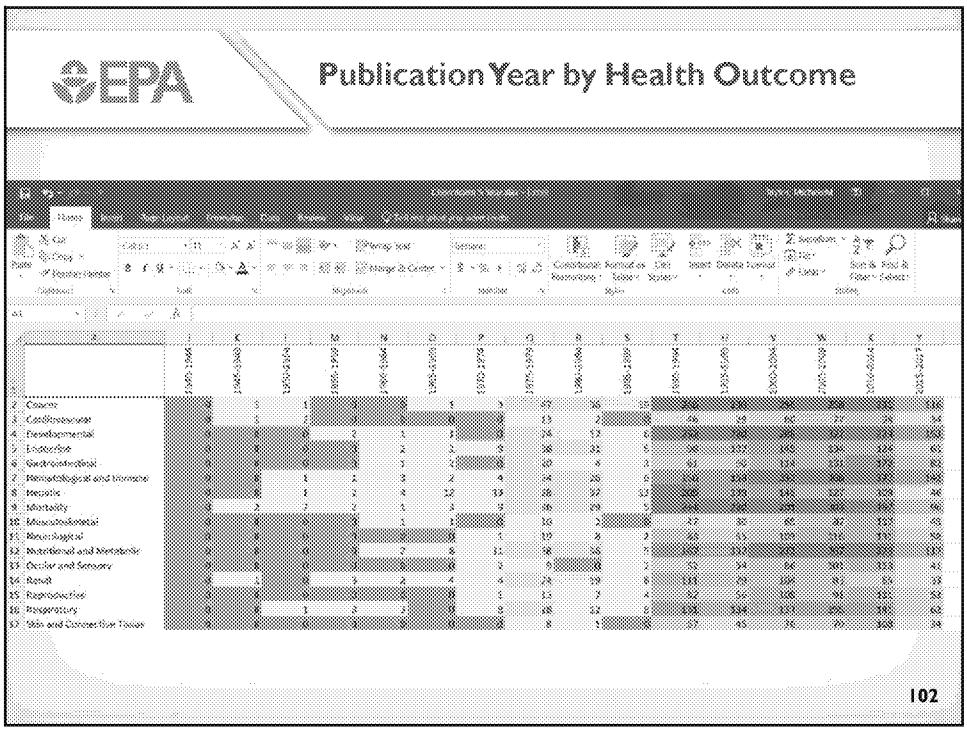


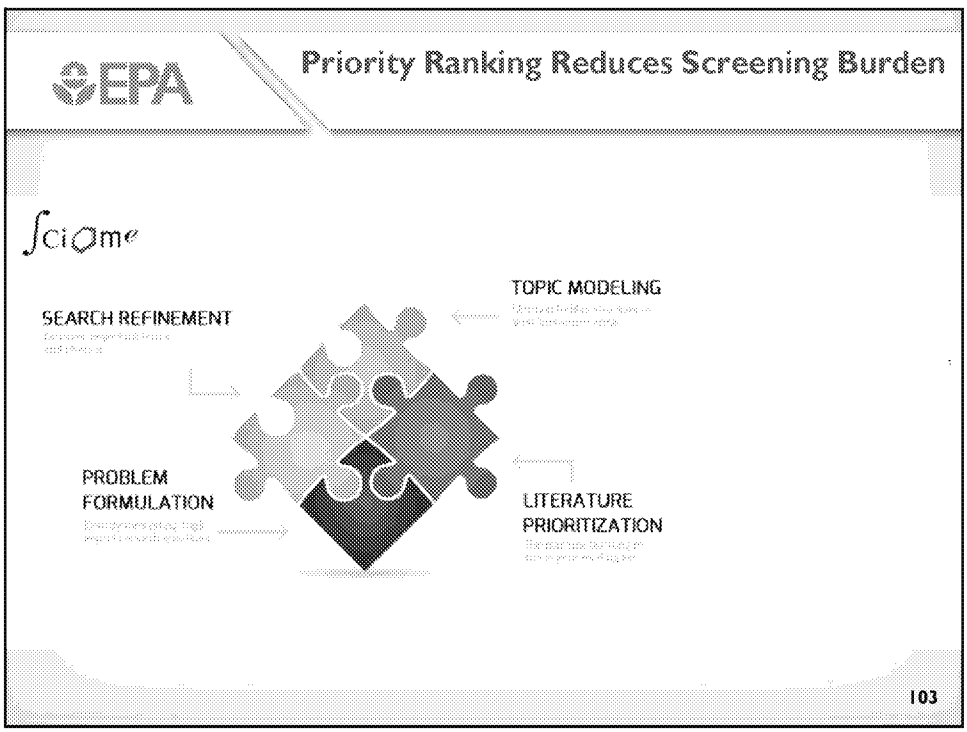
Appendix C

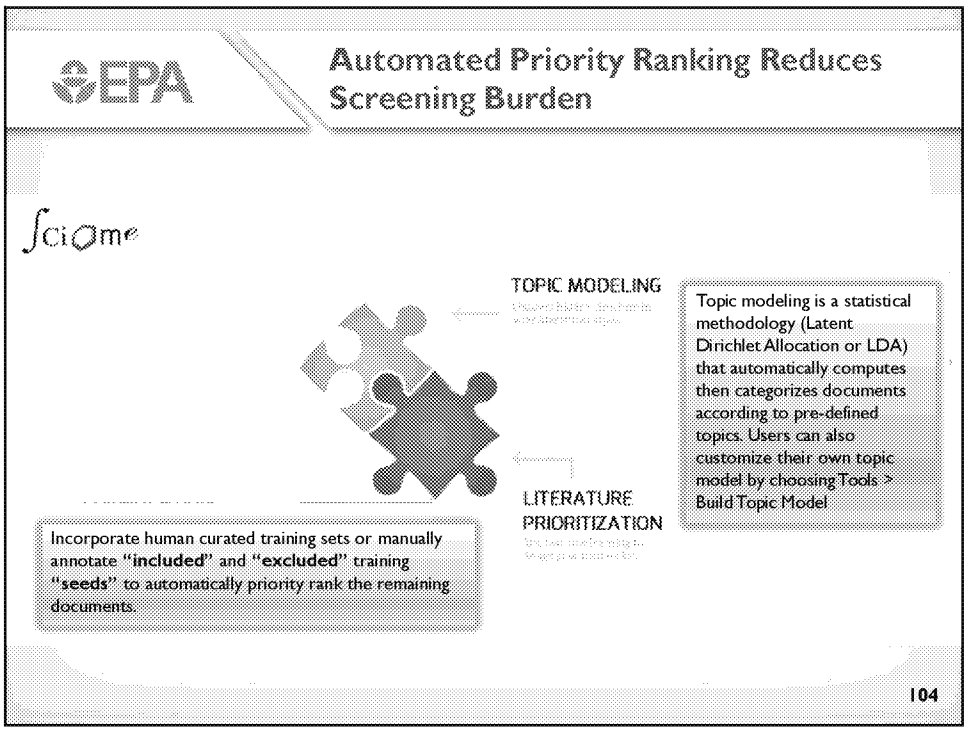


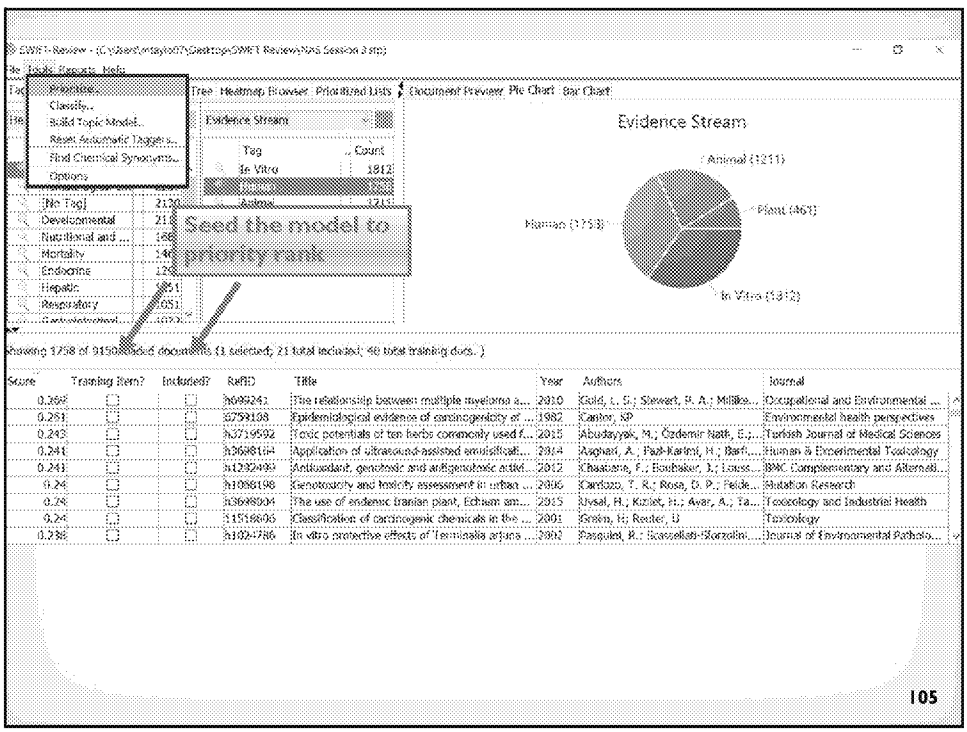


Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

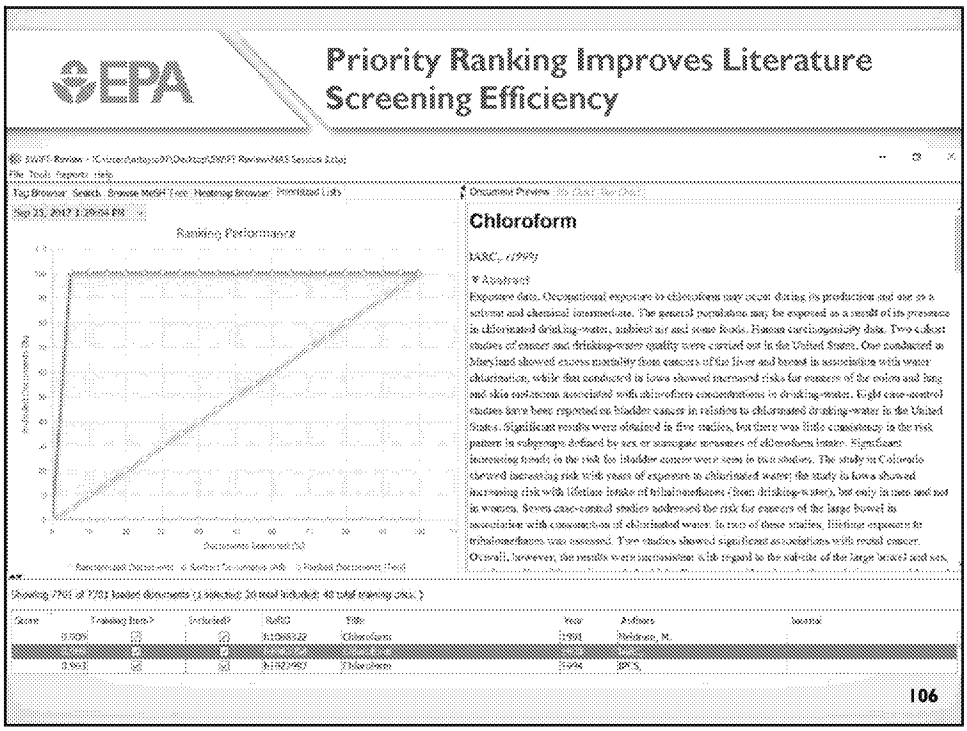


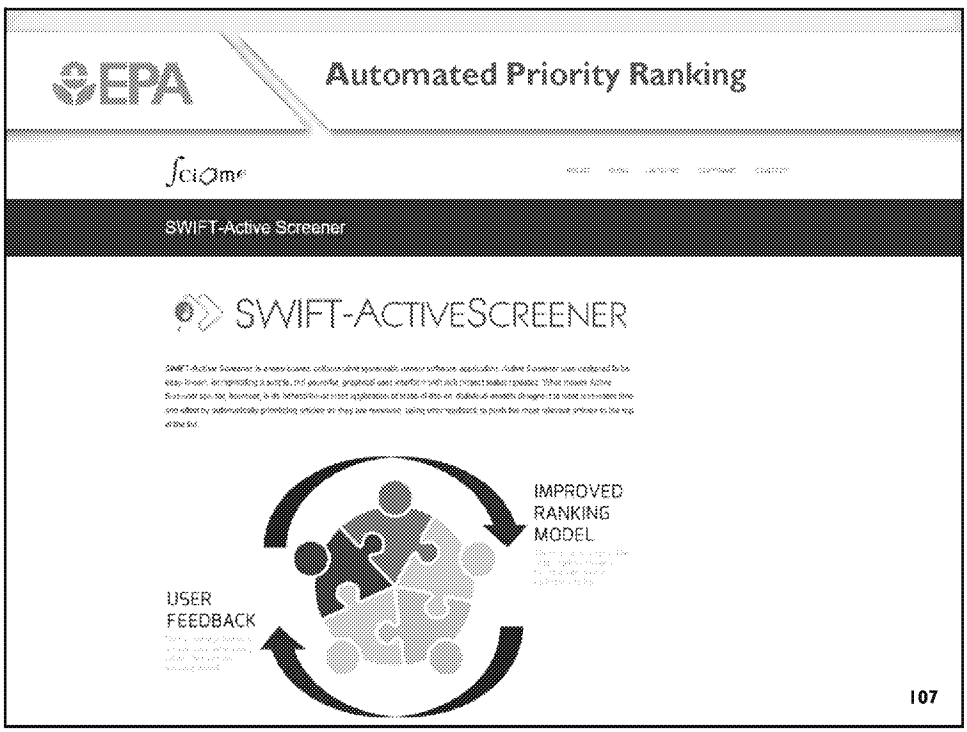






Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation



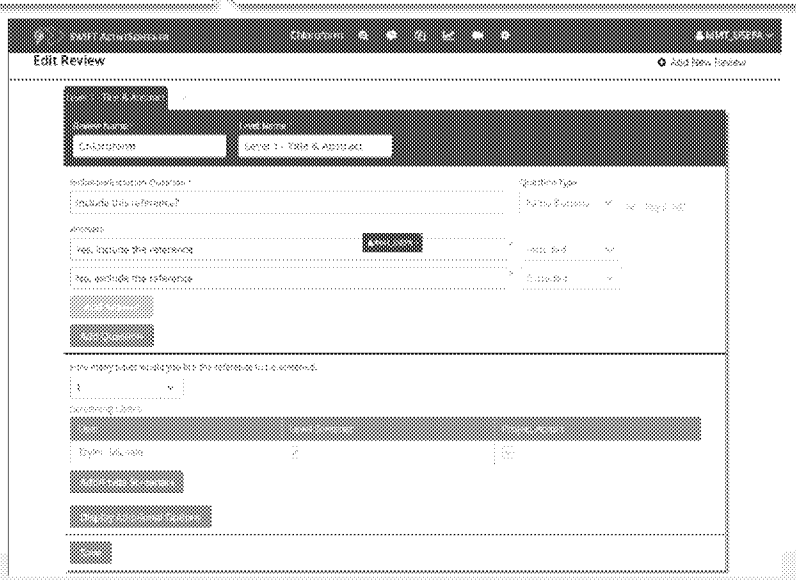


SWIFT Active Screener Capabilities - Improved Ranking Model

- **Web-based, real-time, collaborative, systematic review software application**
- **State-of-the-art statistical models prioritize articles *as they are being reviewed***
- **Experience suggests screening burden is reduced by at least 50% (likely more)**
- **Algorithm improves from screener-input without training "seeds" further increasing efficiency (more efficient than implementing a "seed studies" only model)**
- **Option to "seed" studies if relevant on/off topic literature has been identified**
- **Incorporates a graphical user interface to provide project status updates**
- **User-defined screening levels**
  - **Level 1: Title and Abstract**
  - **Level 2: Full text screening**
  - **Level 3: Conflict Resolution**

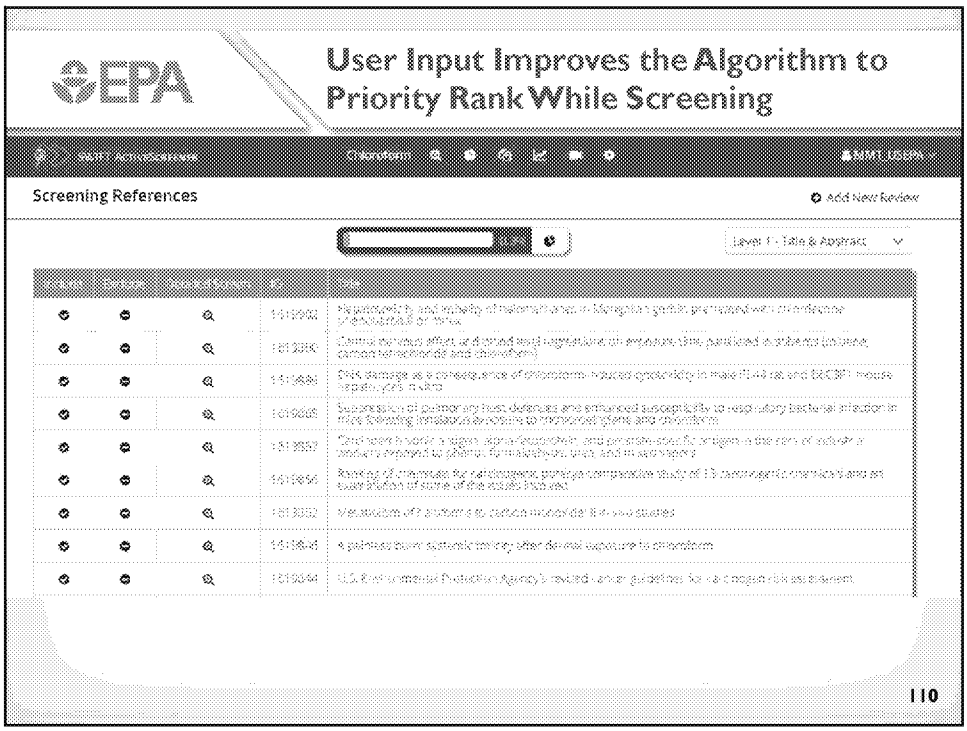
108

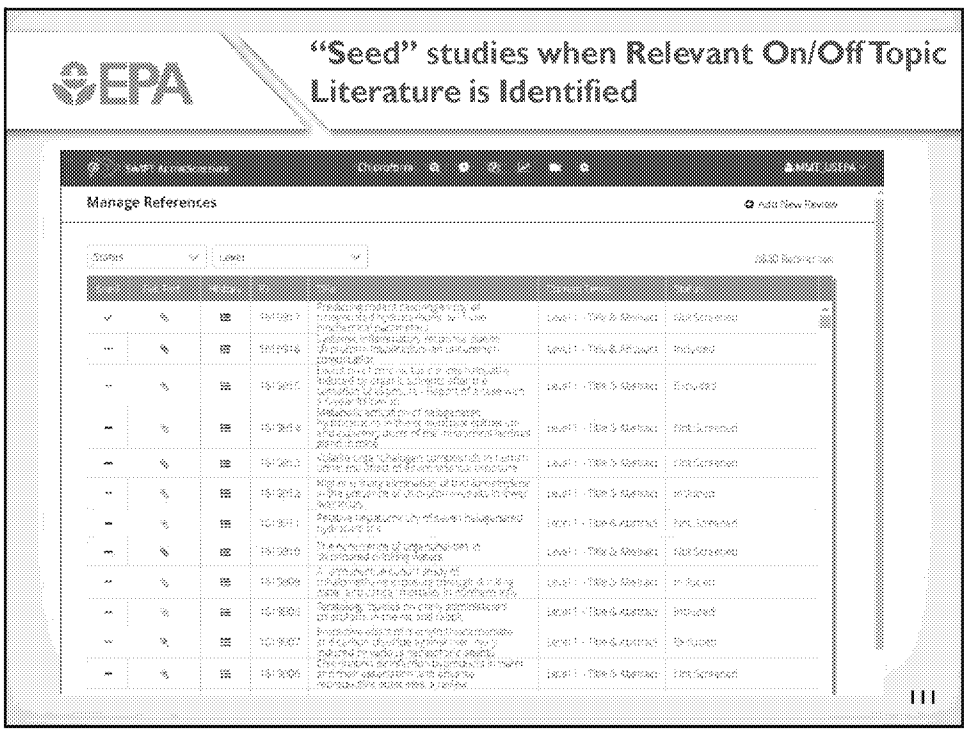
Customize Inclusion/Exclusion Criteria According to the PECO Statement



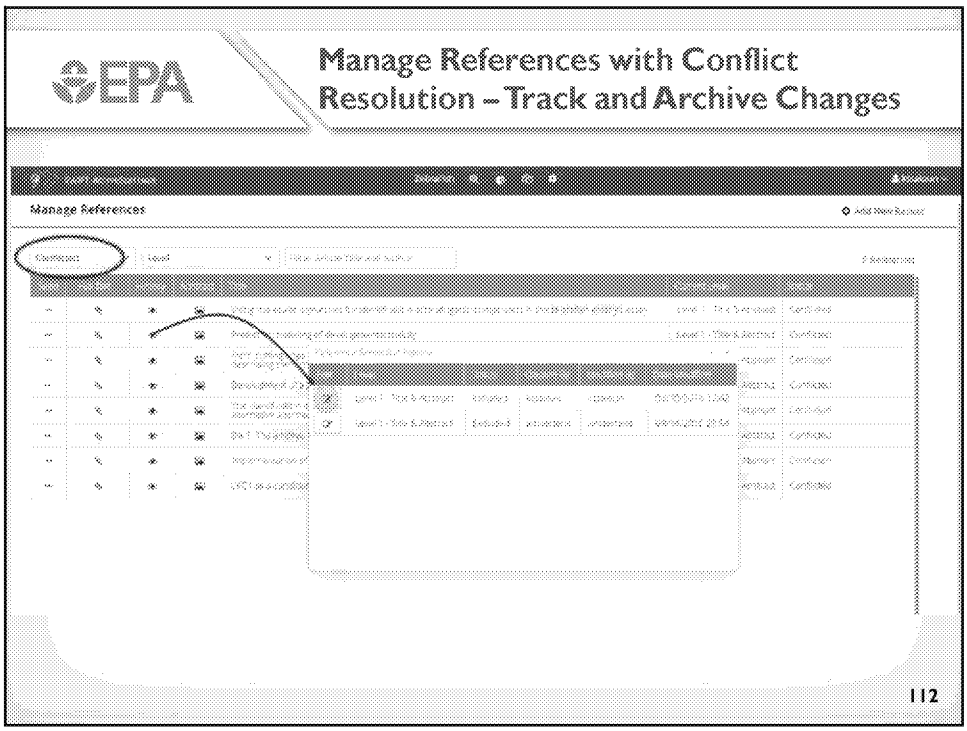
109

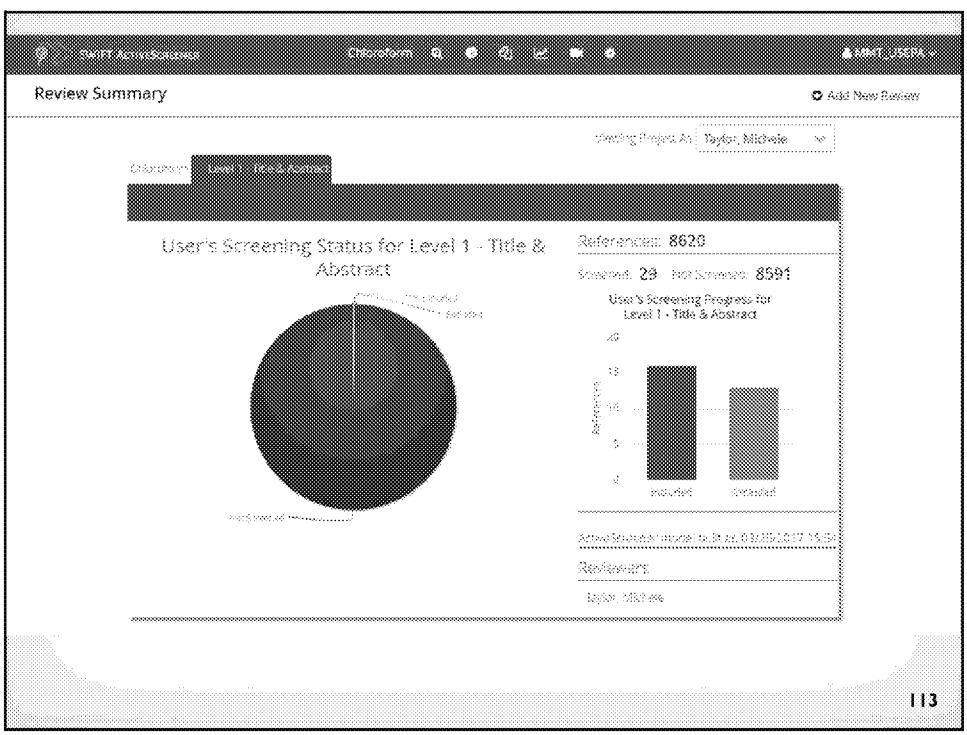
Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation





Appendix C





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

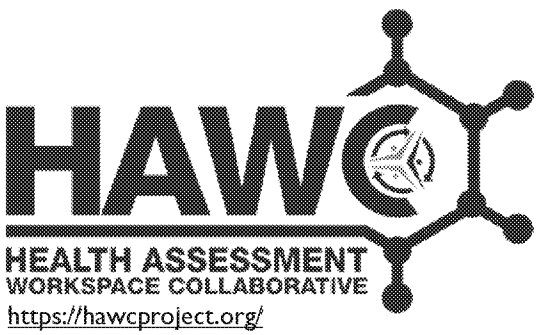
SWIFT Active: Data Integration

- **Active Screener integrates with systematic review tools already in use:**
  - Accepts imports from bibliographic databases and reference curation platforms including *SWIFT Review*, EndNote, Mendeley, Zotaro, and PubMed
  - Results from screening in Active Screener can be exported in standard data formats compatible with applications including *HAWC* and Excel, EndNote, Mendeley, and Zotaro

Current Users

114

HAWC: Study Evaluation, Extraction, Visualization and Data Sharing



https://hawcproject.org/

115

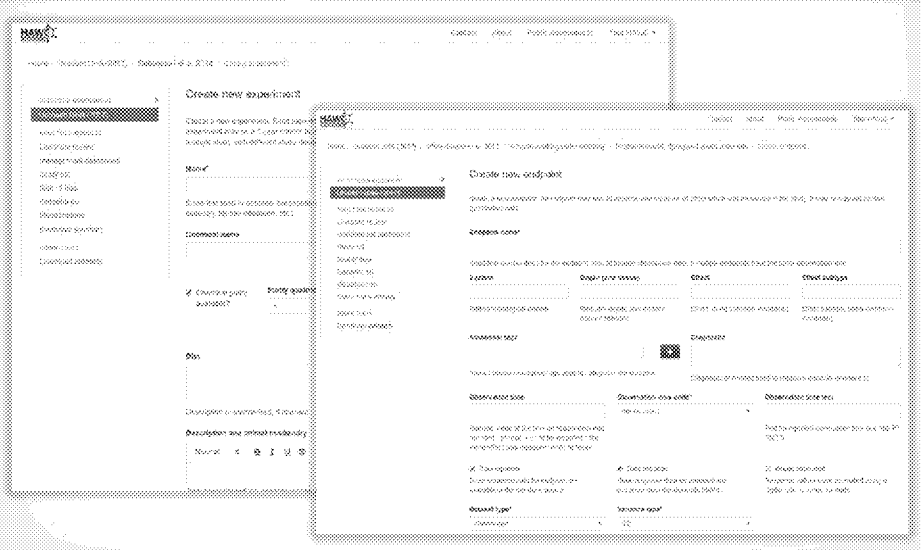
HAWC Capabilities

- **Free and open source**
- **Developed at UNC by Andy Shapiro* with Ivan Rusyn**
- **Literature search and initial screening**
- **Animal bioassay, epidemiological, and in vitro structured study methods/data extraction and visualization**
- **Interactive "click to see more" graphics**
- **Risk of bias and sensitivity evaluation**
- **Modular to work with other tools and maximize flexibility for users**
- **Works best in Google Chrome (preferred), Mozilla Firefox, and Safari**

*current affiliation is National Institute of Environmental Health Sciences/National Toxicology Program (NIEHS/NTP)

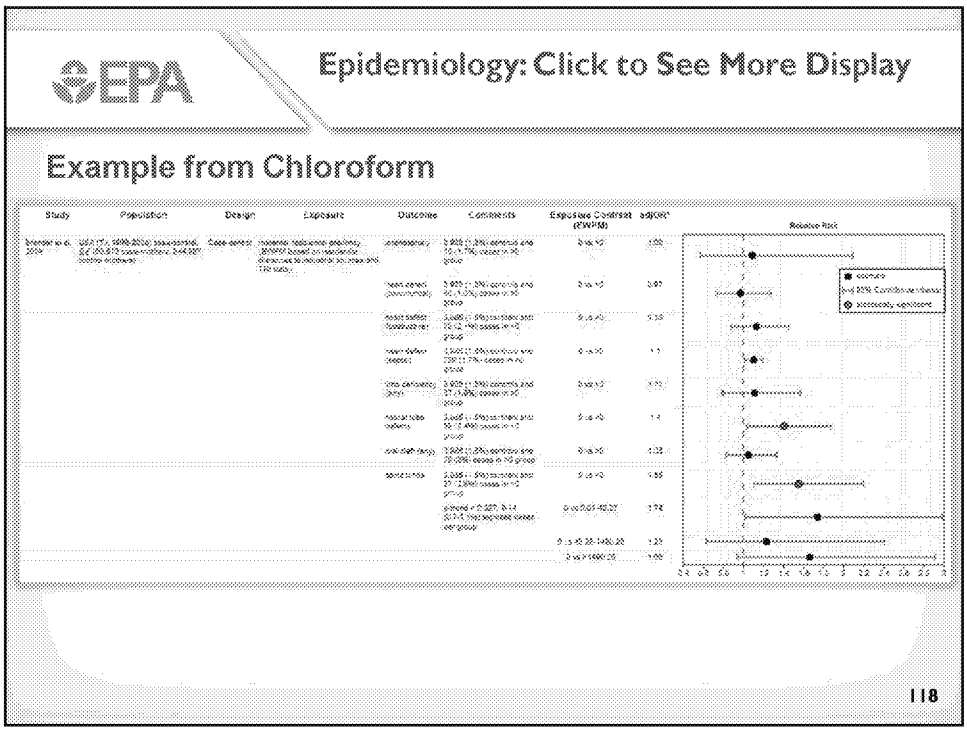
116

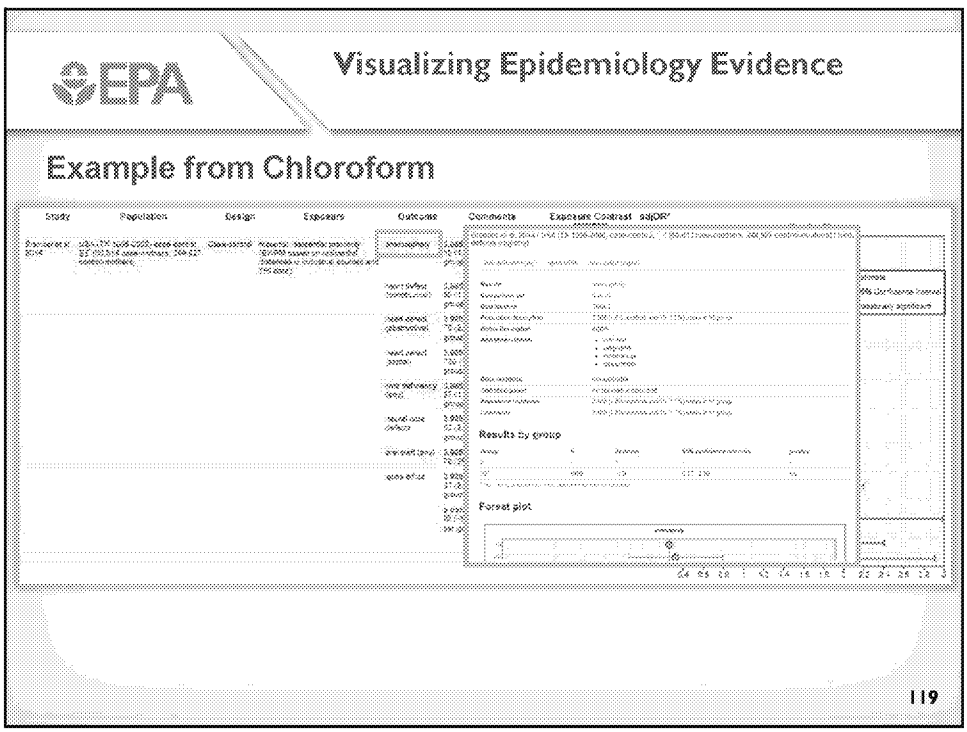
HAWC: Summarizing Animal Bioassays

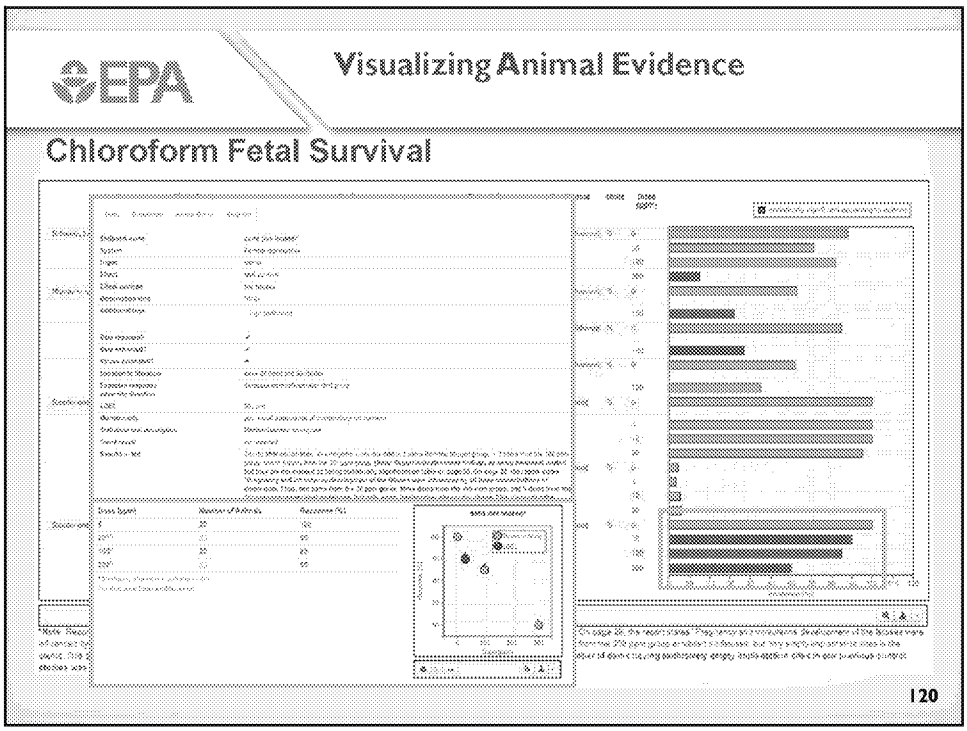


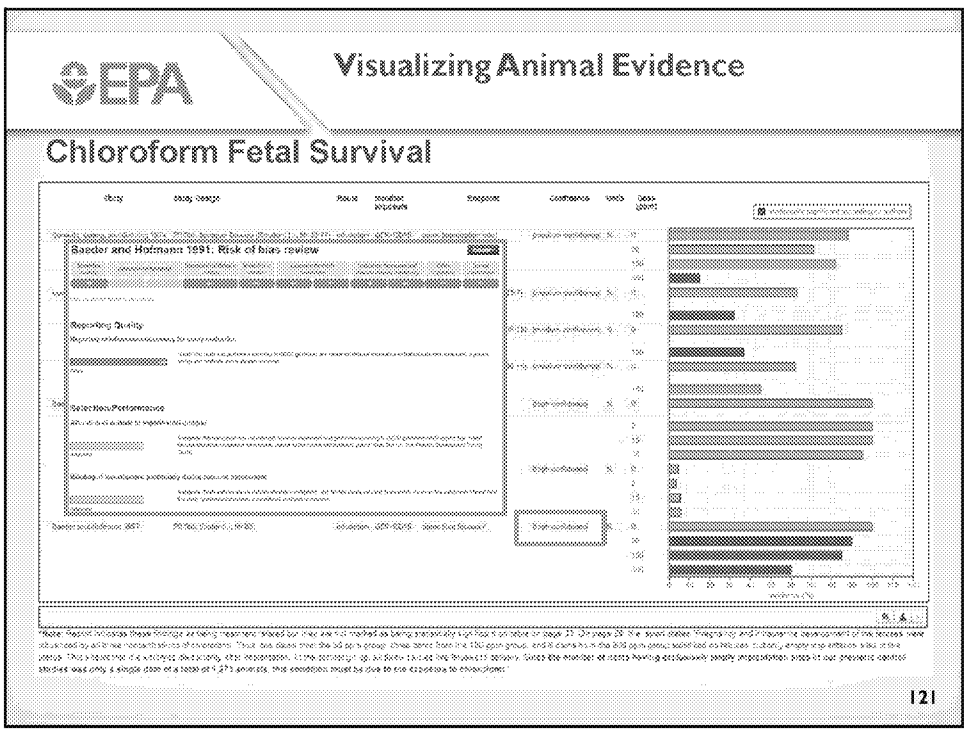
117

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

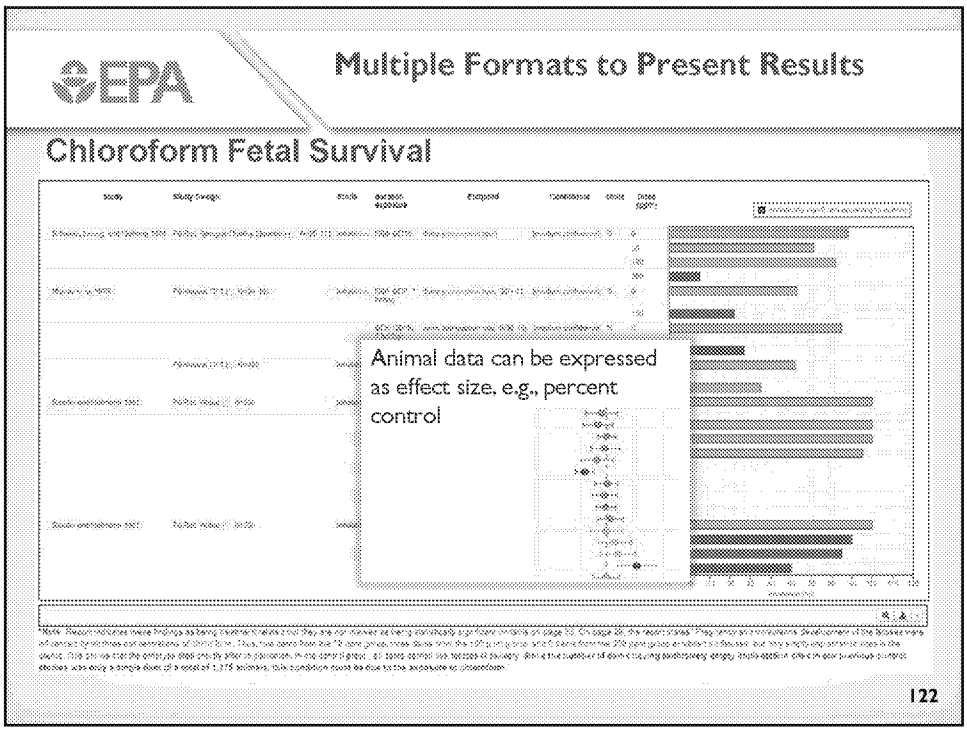


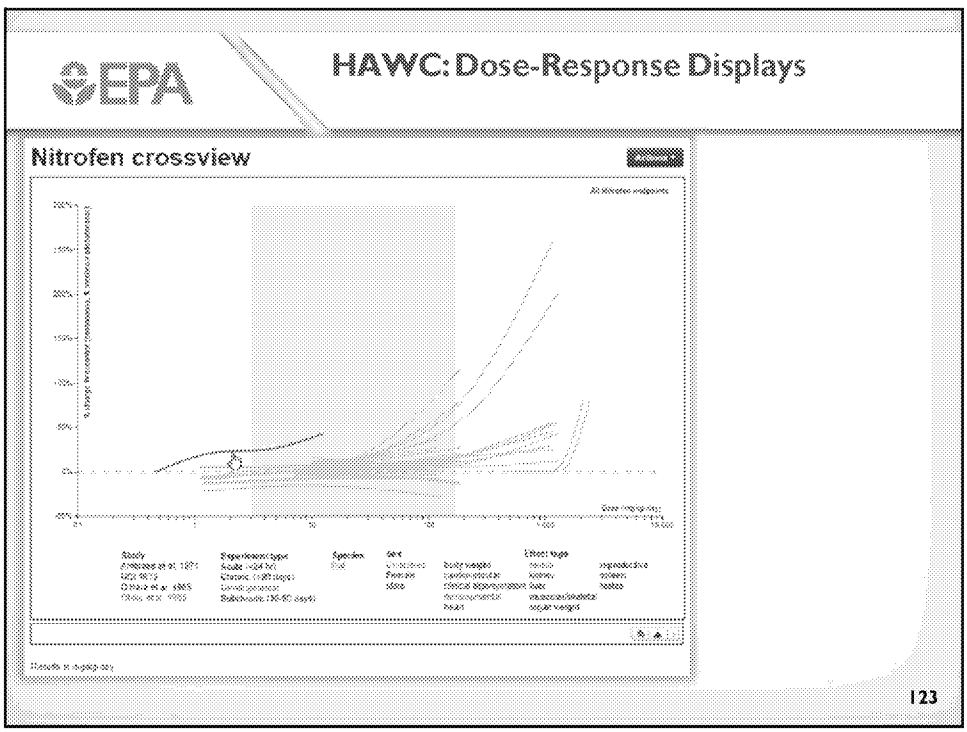


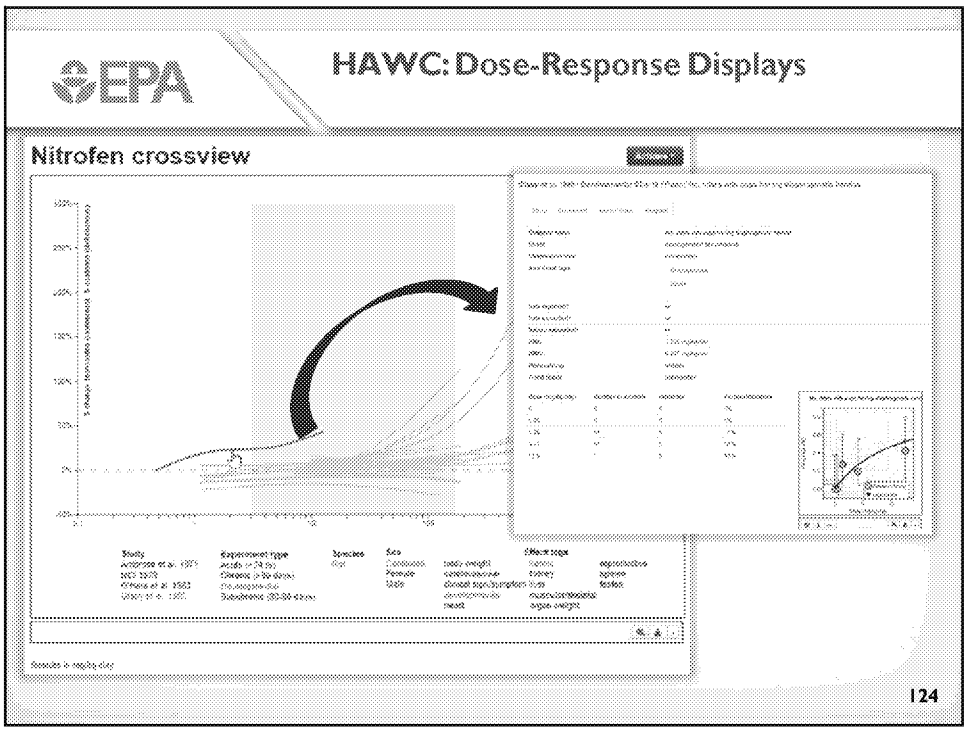


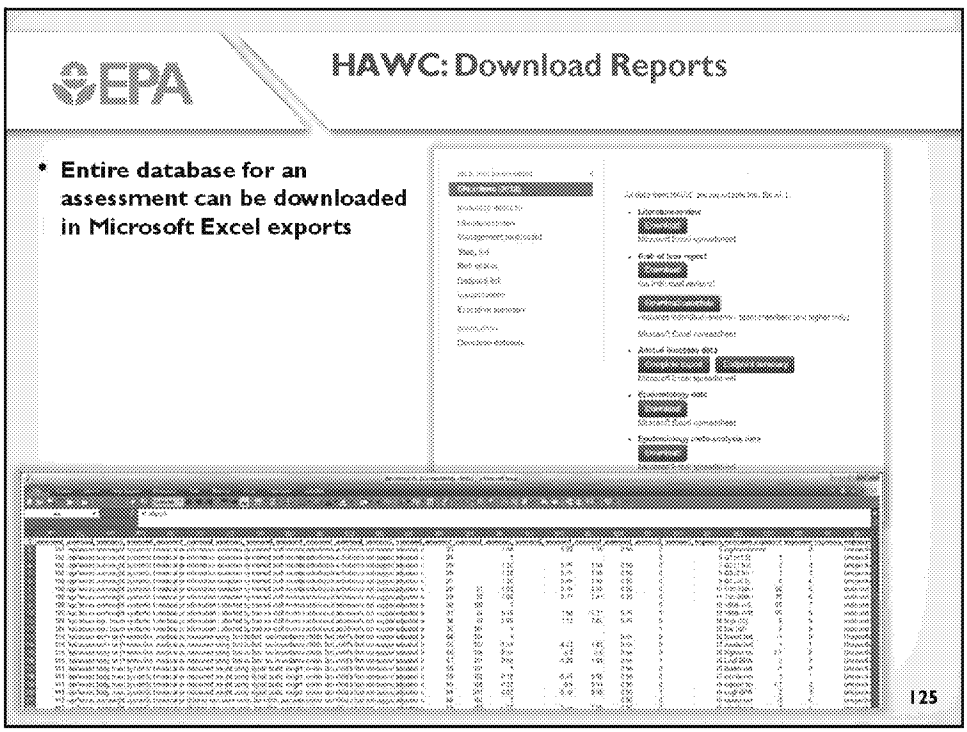


Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

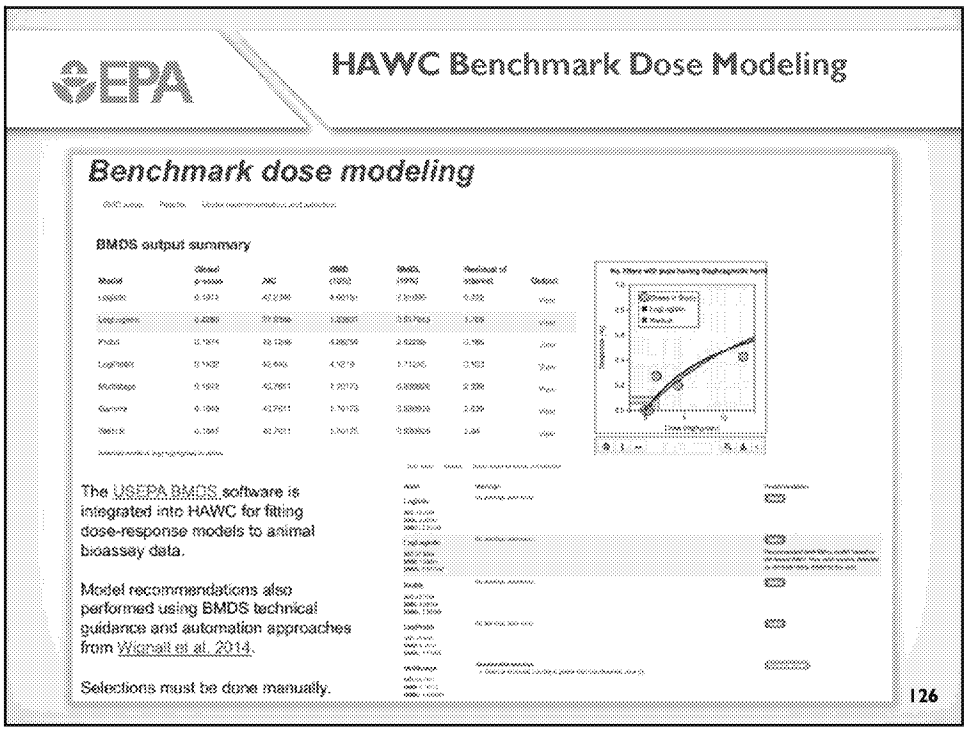


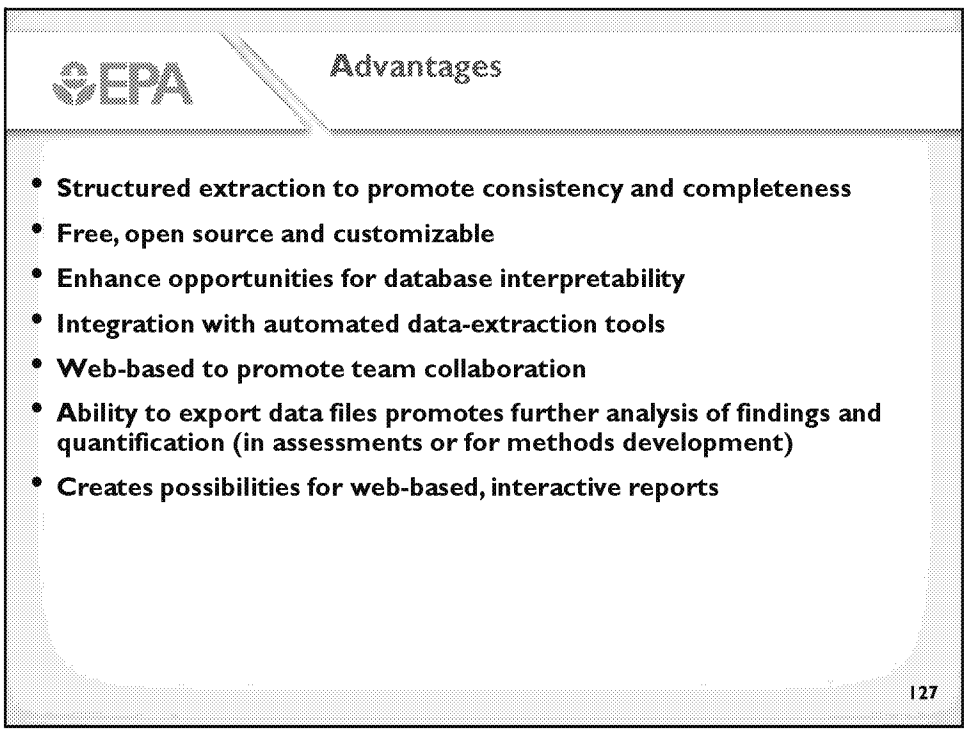






Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation





IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
Looking Forward	• Specialized software tools for efficiency and more user friendly and transparent formats for evidence display have been adopted • Strategic planning on use of text and data-mining tools and automation • Specialized tools facilitate transparent documentation, consistency across assessments, and database interoperability

See Demonstrations:

- SWIFT Review and SWIFT Active
- Health Assessment Workspace Collaborative
- Heath Effects Research Online

128

SESSION 4: STUDY SELECTION FOR DEVELOPING TOXICITY VALUES, AND ADVANCING RESEARCH ON QUANTITATIVE ANALYSES FOR EVIDENCE INTEGRATION AND DOSE-RESPONSE ANALYSES

David Bussard*, Jason Lambert*, Ted Berner, Allen Davis, Jeff Gift, Karen Hogan, Leonid Kopylev, Ravi Subramaniam

[*Speaking]

Office of Research and Development
NCEA, IRIS

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

NAS 2014: Three High Priority (Box 8-1) Recommendations on Quantification

- **TOXICITY VALUES: "EPA should develop criteria for determining when evidence is sufficient to derive toxicity values."**
  - Overall hazard conclusions inform decision whether to develop toxicity values.
  - Better documenting considerations on which studies are carried forward to dose-response.
- **POINTS OF DEPARTURE (PODs): "EPA should clearly present two dose-response estimates: a central estimate (such as a maximum likelihood estimate or a posterior mean) and a lower-bound estimate for a POD from which a toxicity value is derived."**
  - Central estimates (MLEs) of BMDs provided in IRIS assessments along with BMDLs.
  - Will start to use WHO/IPCS approach to characterize distributions in final values.
  - Model averaging to characterize model uncertainty.
- **QUANTITATIVE CAPABILITIES: "EPA should expand its ability to perform quantitative modeling of evidence integration; in particular, it should develop the capacity to do Bayesian modeling of chemical hazards. ...The Committee emphasizes that... IRIS assessments should not be delayed while this capacity is being developed."**
  - Meta-analysis of human and animal studies increasing: hazard decisions and dose-response.
  - Bayesian methods are being explored to help characterize uncertainty.
  - New approach methods and assays are increasingly being evaluated quantitatively.

130

Evidence Integration Conclusions Inform when to Develop Toxicity Values

Evidence integration conclusion	Quantitative toxicity value provided?
Strongest conclusion for a human health effect (for cancer, a descriptor of *Known*)	Yes.
Moderately strong conclusion for a human health effect (for cancer, a descriptor of *Likely*)	Yes.
Weakest conclusion for a human health effect (for cancer, a descriptor of *Suggestive*)	Determined by situation (e.g., may provide values when useful for decision purpose and the evidence includes a well-conducted study)
Inadequate information	No, although bounding estimate from a study that does not show positive results can be derived where useful for decision purpose.
Strong support for no human health effect	No.

131

Decision-Making for Advancing Studies to Develop Toxicity Values

IRIS has further clarified the considerations that inform the selection of studies to estimate human dose-response relationships (next slide).

- IRIS continues to find that this decision process is not reducible to a formula.
- Expert judgment is essential for judging the relative merits of individual studies and which studies support more integrative quantitative analyses (e.g., meta-analysis).
- IRIS must often utilize studies with a range of attributes and levels of reporting. For example, the available studies on many mission-critical chemicals do not provide data on an individual subject basis.
- For full transparency, IRIS continues to emphasize documentation of the factors it weighed in emphasizing certain studies, or combinations of studies, over others.

132

More Explicitly Defining the Attributes IRIS Uses to Evaluate Studies for Derivation of Toxicity Values

In addition to qualitative study evaluation judgments (i.e., *medium* or *high* confidence studies are preferred), studies are assessed across several study attributes

Example Primary Considerations for Selection of Studies for Derivation of Toxicity Values			
Study attribute		**Human studies**	**Animal studies**
Test species		**Human data are generally preferred** to eliminate interspecies extrapolation uncertainties (e.g., in toxicodynamics and specific health outcomes).	**Animals that respond most like humans are preferred**. Outcomes associated with species known to show differences in sensitivity can provide support with suitable qualification.
Human relevance of the exposure paradigm	**Exposure route**	Studies involving **typical human environmental exposure routes are preferred** (e.g., oral, inhalation). A validated toxicokinetic model can be used to extrapolate across exposure routes.	
	Exposure duration	For chronic toxicity values, **chronic or subchronic studies are preferred. Exceptions exist** (e.g., when a population or lifestage is more sensitive during a particular time window)	
	Exposure levels	**Exposures near the range of typical environmental human exposures are preferred. Studies with a broad exposure range and multiple exposure levels are preferred** to the extent that they can provide information about the shape of the exposure-response relationship* and facilitate extrapolation to more relevant (generally lower) exposures.	
Susceptibility		Studies that yield **risk estimates in the most susceptible groups are preferred.** Inclusion of design features in the analysis (e.g., matching procedures, blocking; covariates or other procedures for statistical adjustment) **that adequately address the relevant sources of potential critical confounding for a given outcome are preferred.**	

*U.S. EPA Benchmark Dose Technical Guidance (2012)

133

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

IRIS Assessments Are Providing Central MLE Estimates of BMDs Along with BMDLs

Table 2-1. Summary of derivation of PODs

Endpoint and reference	Species/ sex	Model[a]	BMR	BMD mg/kg-d	BMDL mg/kg-d	POD mg/kg-d	POD$_{adj}$[b] mg/kg-d	POD$_{HED}$[c] mg/kg-d
Developmental								
Neurobehavioral changes: Open field crossed squares at PND 69 (Chen et al. (2012))	Male and Female Sprague-Dawley rats	Exponential 4	1 SD	0.23	0.13			
Neurobehavioral changes: Elevated plus maze open arm entries at PND 70 (Chen et al. (2012))	Female Sprague-Dawley rats	Exponential 4	1 SD	0.21	0.092	0.092[d]	0.092	0.092
Neurobehavioral changes: Morris water maze hidden platform trial escape latency at PNDs 72–74 (Chen et al. (2012))	Male and Female Sprague-Dawley rats	Hill CV Hill CV Hill CV Hill NCV	1 SD (5 sec)	PND73: 0.69 PND73: 0.33 PND75: 0.27 PND74: 0.23	0.18 0.16 0.12 0.15			
Cardiovascular effects at PND 63 (Jules et al. (2012))	Long-Evans rats	LOAEL (0.8 mg/kg-d) (15% ↑ in systolic blood pressure; 12% ↑ in diastolic blood pressure)				0.8	0.8	0.23

Recent animal study example to the left: **Benzo[a]pyrene (EPA, 2017)**

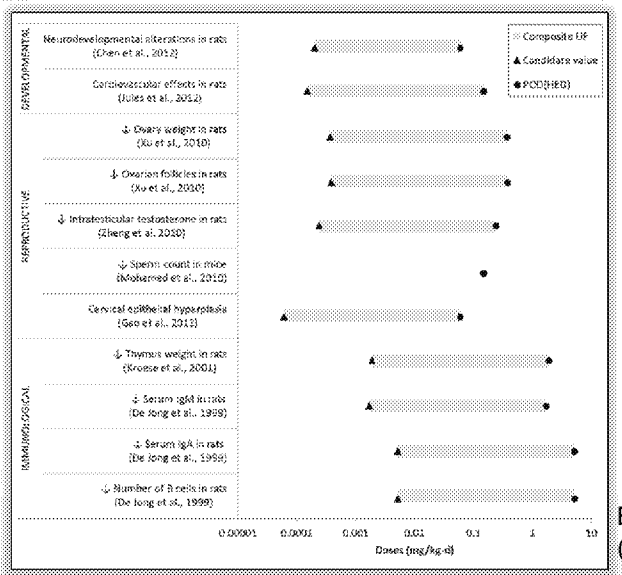
p. 2-8 Toxicological Review of Benzo[a]pyrene https://cfpub.epa.gov/ncea/iris_drafts/documents/documents/toxreviews/0136tr.pdf

Recent epidemiology example: **Ethylene oxide (EPA, 2016)**

p. 4-109 Toxicological Review of the Inhalation Carcinogenicity of Ethylene Oxide. (EPA, 2016) https://cfpub.epa.gov/ncea/iris_drafts/documents/documents/toxreviews/1021tr.pdf

134

IRIS is also Presenting Arrays of Candidate Toxicity Values



Benzo[a]pyrene (EPA, 2017)

135

EPA

Improvements in Characterizing Uncertainty

1) Model Averaging: characterizing model uncertainty

- Currently evaluating several methods
- Approach for dichotomous data expected to undergo peer review in 2018

$$\Pr(BMD \mid D) = \sum_{i=1}^{9} \pi_i \Pr(BMD \mid M_i, D)$$

Posterior Distribution of the BMD

$$\alpha = \int_{-\infty}^{BMD_\alpha} \Pr(BMD \mid D) dBMD$$

Calculation of the BMDL

136

EPA

Improvements in Characterizing Uncertainty

2) Distributions and Central Estimates: characterizing uncertainty in the human toxicity value

- WHO/IPCS guidance (IPCS, 2014)
- Risk-specific doses in terms of ranges, explicitly described:
  - Effect magnitudes
  - Confidence levels
  - Human population incidence rates.
- A probabilistic approach to adjustments from animal to human; a framework for refining toxicity values.

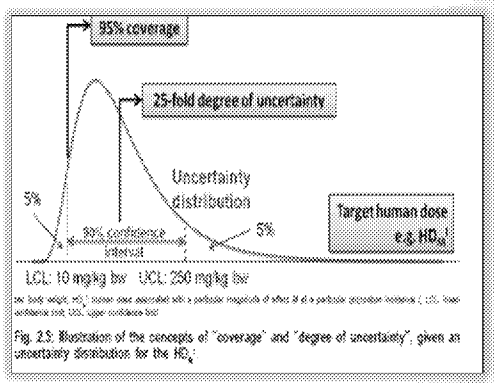


Fig. 2.5. Illustration of the concepts of "coverage" and "degree of uncertainty", given an uncertainty distribution for the HD_5

137

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

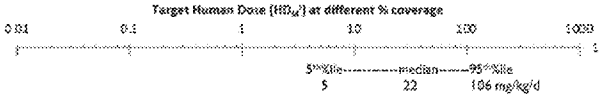
Improvements in Characterizing Uncertainty

WHO/IPCS Approach:

IRIS intends to provide such calculations along with traditional Reference Values:

- Confidence intervals on risk-specific doses
- Central estimates
- Estimates of incidence as a function of dose
- Use of appropriate probability math for uncertainty adjustments (instead of UFs) to allow for a more probabilistic and scientific value for use in risk assessment

By characterizing ranges of risk-specific doses, this provides more than a "conservative" estimate (it provides useful context by estimating the full distribution)



138

Use of Quantitative Modeling to Inform Evidence Integration

Meta-Analysis:

Increasingly Being Used to Interpret Sets of Results across Similar Populations

- Formal tools continue to be used to combine similar human epidemiology studies to improve decisions about hazard and about slope of dose-response.
- These approaches have also been used to better understand animal data that differ between studies of similar species and endpoints.
- As software tools and best practices become more common and easier to apply to environmental health studies, IRIS intends to consider their use more routinely.

Other examples: Libby Amphibole Asbestos (2014) and Trimethylbenzene analysis (Davis and Kraft, 2017) – see poster session; Arsenic assessment (in process)

139

Use of Quantitative Modeling to Inform Evidence Integration

Bayesian Approaches:

More Frequent Use Across Different Applications, and Research is Ongoing

- **Characterizing Uncertainty**
  - Bayesian approaches were used to characterize uncertainty in PBPK modeling and evaluate inter-related model inputs (Perchlorate peer review, 2018).
  - Bayesian Analysis is compatible with the WHO/IPCS Approach for characterizing uncertainty
- **Model Averaging**
  - Bayesian approaches are being applied to individual BMD models, and then model averaging is used to characterize uncertainty
- **Meta-Analysis**
  - Bayesian meta-analysis is currently being used to evaluate arsenic epidemiology studies
- **Bayesian Networks** (exploratory research is currently underway)
  - Possess the potential to integrate across evidence streams and bridge data gaps, borrowing strength from diverse data.
  - Software and mathematics are currently available.

140

Future work to better meet Agency needs for "benefits analysis"

Economics benefits analysis would ideally estimate incidence resulting from different decision options.

- We have provided human dose response functions from some analyses based on epidemiology data. (Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide, EPA, 2016).

IRIS is also evaluating analogous predictions from animal data that could inform benefits analysis, including modifications of the IPCS approach.

141

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Advancing Application of New Approach Methods (NAM) and Data in HHRA

- Over the past decade, several reports, books, resource documents, etc. have been published regarding the use of New Approach Methods (NAM) across the human health risk assessment paradigm (i.e., shifting the paradigm)
- Numerous labs, centers, workgroups, and initiatives across federal, private, and academic institutions have been formed to advance NAM

- EPA/ORD/NCEA, in conjunction with partners (e.g., NCCT, NTP) has been actively engaged in the conceptualization and evaluation of NAM across a broad landscape of HHRA applications

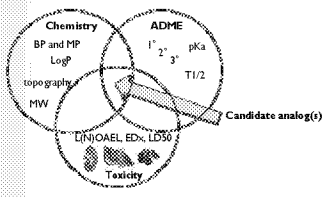
142

NAM Toolbox to Date

- **Data-mining**: ToxRefDB-comprehensive collection and collation of extant hazard and exposure data –(Martin et al. 2009. Env Health Perspect 117: 392-399)
- **Chemoinformatics**: structure-activity/read-across; QSAR –(Wang et al. 2012. Regul Toxicol Pharmacol 63: 10-19; Craig et al. 2014. J Appl Toxicol 34: 787-794)
- **High-Throughput (HT) Exposure modeling**: ExpoCast –(Egeghy et al. 2016. Env Health Perspect. 124(6):697-702)
- **HT Toxicokinetics**: in vitro to in vivo (IVIVE) modeled dosimetry –(Wambaugh et al. 2015. Tox Sci 147: 55-67)
- **Bioactivity**: short-term animal; cell-free and/or cell-based HT assay data – (Judson et al. 2011. Chem Res Toxicol 24: 451-462; Dean et al. 2017. Tox Sci 157(1):85-99)
- **Adverse Outcome Pathway (AOP)**: expert-driven identification of signal transduction pathways along the exposure to outcome continuum. –(Edwards et al. 2016. J Pharmacol Exp Ther. 356(1):170-181)

143

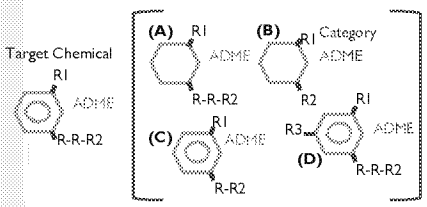
Chemoinformatics



Expert-driven Read-Across

Data-poor chemicals
- Inferred/interpolated hazard
- Surrogate based POD and subsequent derivation of RfVs

IRIS-type chemicals
- Data-gap filling
- Augment WOE
- Potential for reducing uncertainties



ADME = Absorption, Distribution, Metabolism, Elimination

Category approach

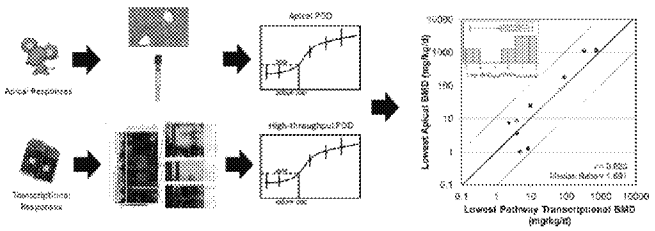
Data-poor chemicals
- Data-gap filling
- Extrapolated hazard
- Less applicable for quantitative assessment currently

IRIS-type chemicals
- Data-gap filling
- Augment WOE
- Foundational member of category (i.e., anchor chem)

- Similarity in structure and physicochemical properties between a chem of concern and a population of analogs
- Robustness of approach dependent on density of analogs populating a category
- Highly reliant on WOE supporting toxicity endpoints across category
- Presumes common Adverse Outcome Pathway or Mode of Action across category members

144

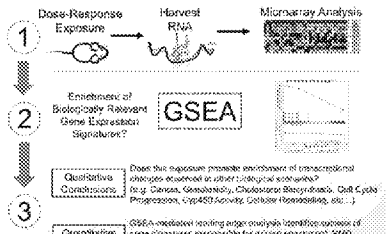
Bioactivity (e.g., transcriptomics)



- Close relationship between genotype/phenotype across two different routes of exposure, rodent species, and multiple target tissues
- In vitro?? Will need to optimize metabolism protocols; integrate IVIVE

Data-poor chemicals
- Evidence base for hazard
- Empirical dose-response based on pathway perturbations
- Reduce need for longer-term animal studies

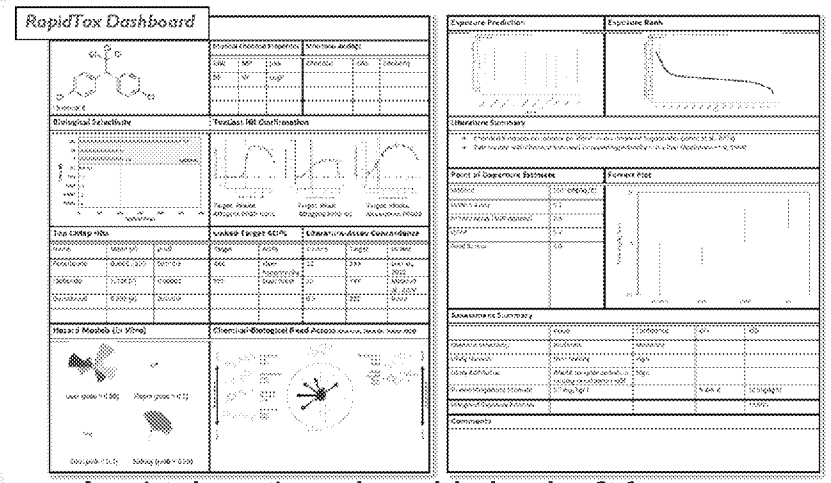
IRIS-type chemicals
- Augment WOE (e.g., MOA/AOP)
- Opportunity to alert off-target effects
- Potential for reducing uncertainties



145

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Integrated Application to Risk Assessment



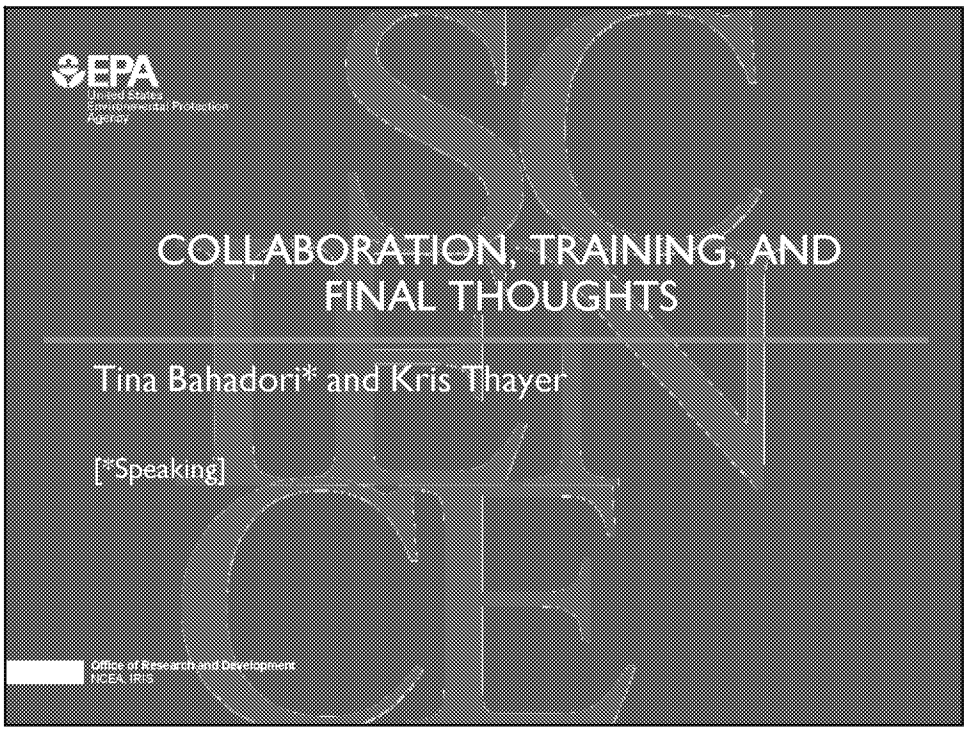
- **Associated narrative can be modular based on fit-for-purpose**
- **Systematic WOE always, but can be graded based on decision context**
- **Characterization of qualitative and quantitative uncertainties**

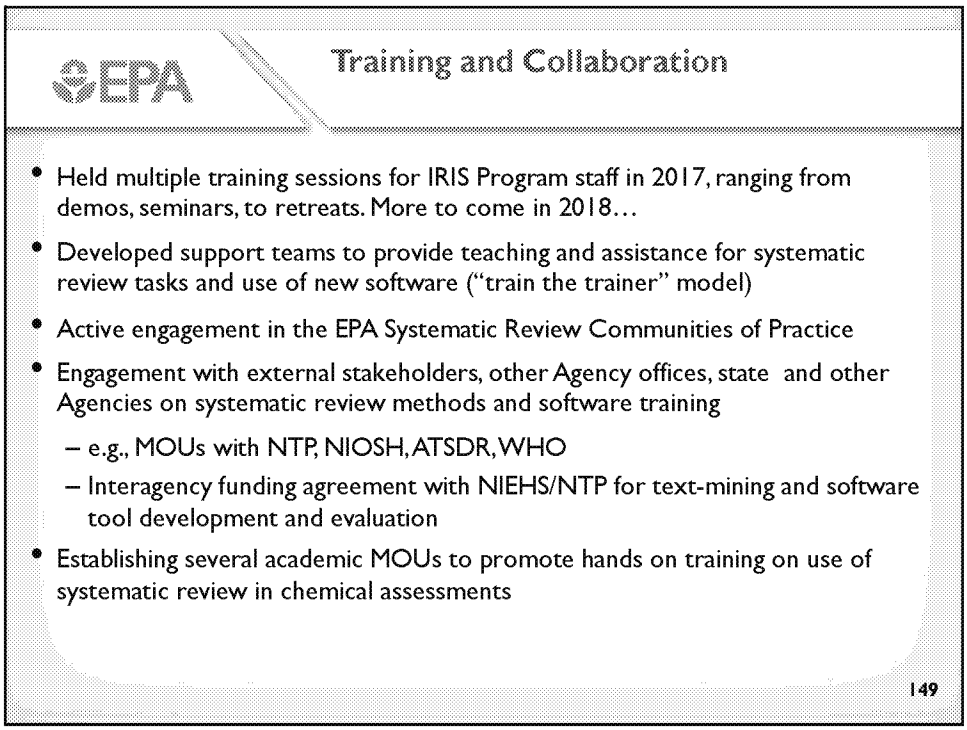
146

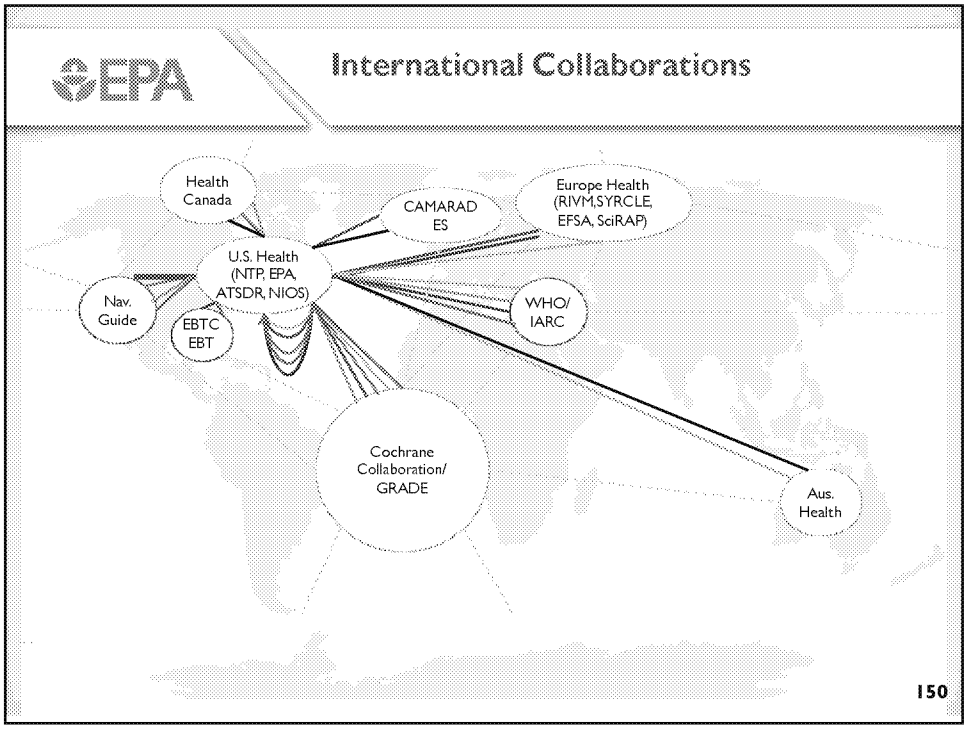
IRIS has Addressed the Major NAS 2014 Recommendations

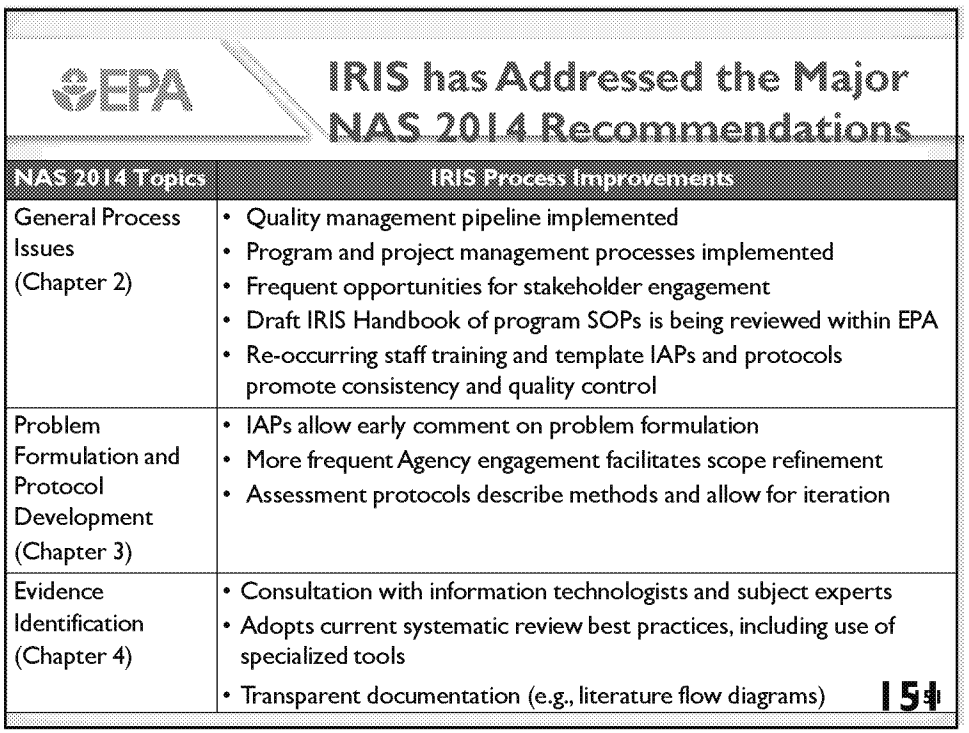
NAS 2014 Topics	IRIS Process Improvements
Evidence Integration for Hazard Identification (Chapter 6) and Derivation of Toxicity Values (Chapter 7)	• Developing and applying quantitative tools in support of evidence synthesis and integration, including meta-analytical approaches • Expanded development and use of more advanced quantitative methods in software tools, such as BMDS • Developed more explicit criteria for deriving toxicity values, including the intent to derive quantitative toxicity values when IRIS reaches one of the stronger evidence integration conclusions, as well more specific criteria for the evaluation of individual studies • Providing MLE estimates of BMDs, along with BMDLs • Applying and exploring quantitative approaches to better characterize uncertainty, including probabilistic and Bayesian approaches
Future Directions (Chapter 8 "Lessons Learned" and "Looking Forward")	• Quantitative assessment methods will be updated in a continuing, strategic fashion, including capacity building (e.g., training; evolving best practices) for current approaches including meta-analysis, probabilistic analyses, and Bayesian methods

147









IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
Evidence Evaluation (Chapter 5)	• Individual studies are evaluated for reporting quality, risk of bias, and sensitivity • Decisions and supporting rationale are clearly documented • Study evaluations impact subsequent assessment decisions
Evidence Integration for Hazard Identification (Chapter 6)	• Structured frameworks provide transparency in expert judgments across human, animal, and mechanistic studies (based on Hill) • Standardized templates documenting key evidence integration decisions have been developed (evidence profile tables) • Developing and applying quantitative tools in support of evidence synthesis and integration, including meta-analytical approaches • Expanded development and use of more advanced quantitative methods in software tools, such as BMDS

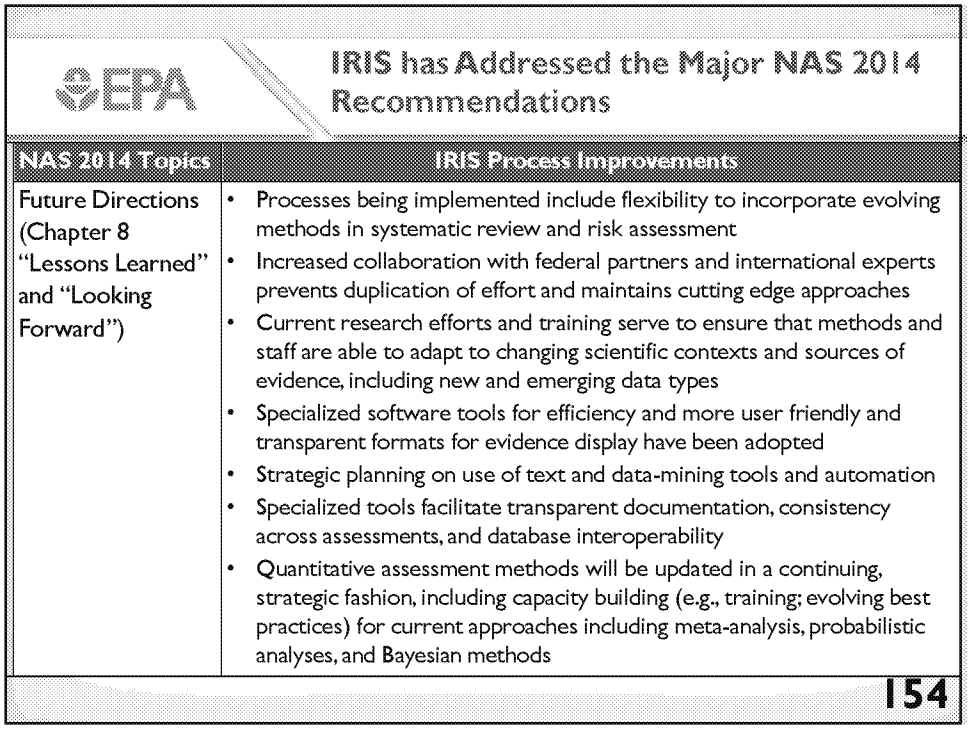
152

IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
Derivation of Toxicity Values (Chapter 7)	• Developed more explicit criteria for deriving toxicity values, including the intent to derive quantitative toxicity values when IRIS reaches one of the stronger evidence integration conclusions, as well more specific criteria for the evaluation of individual studies • Providing MLE estimates of BMDs, along with BMDLs • Applying and exploring quantitative approaches to better characterize uncertainty, including probabilistic and Bayesian approaches

153

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation



EPA

IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
Future Directions (Chapter 8 "Lessons Learned" and "Looking Forward")	• Processes being implemented include flexibility to incorporate evolving methods in systematic review and risk assessment • Increased collaboration with federal partners and international experts prevents duplication of effort and maintains cutting edge approaches • Current research efforts and training serve to ensure that methods and staff are able to adapt to changing scientific contexts and sources of evidence, including new and emerging data types • Specialized software tools for efficiency and more user friendly and transparent formats for evidence display have been adopted • Strategic planning on use of text and data-mining tools and automation • Specialized tools facilitate transparent documentation, consistency across assessments, and database interoperability • Quantitative assessment methods will be updated in a continuing, strategic fashion, including capacity building (e.g., training; evolving best practices) for current approaches including meta-analysis, probabilistic analyses, and Bayesian methods

154

Appendix D

Posters by US Environmental Protection Agency

- D-1: New Approach Methods in Human Health Risk Assessment
- D-2: Combining Data within Species: Meta-analysis in IRIS
- D-3: Systematic Evaluations of Physiologically-Based Pharmacokinetic Models for Human Health Risk Assessment
- D-4: Male Reproductive Toxicity in Animal Studies of Diisobutyl Phthalate (DIBO): A Case Study Application of Systematic Review Approaches
- D-5: Male Reproductive Toxicity in Epidemiology Studies of Phthalates: A Case Study Application of Systematic Review Approaches
- D-6: Quantitative Evaluation of Uncertainty: APROBA and Beyond
- D-7: Mode of Action and Human Relevance Evaluation of Dibutyl Phthalate (DBP)-Induced Male Reproductive System Toxicity
- D-8: EPA Dose-Response & Related Software – New & Future Developments
- D-9: Evidence Profile Table for DIBP and Male Reproductive Toxicity
- D-10: A New Bayesian Approach to Combining Different Species Data

New Approach Methods in Human Health Risk Assessment

Jason C. Lambert
U.S. EPA, Office of Research and Development, National Center for Environmental Assessment - Cincinnati

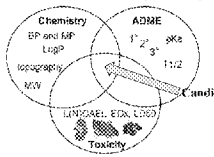
Integration of New Approach Methods-Theory

- Chemicals nominated for Human Health Risk Assessment (HHRA) have widely varying hazard and dose-response databases
- Integration of New Approach Methods (NAM) is therefore fit-for-purpose along a decision-based gradient:
  - Data-poor chemicals⇨ NAM may be a driver
  - Data-rich chemicals ⇨ NAM fills data gaps
  - Same/similar assays, same/similar data can be used in different ways to answer specific questions
- NAMs currently being integrated or evaluated in EPA HHRA contexts include:
  - Read across (expert-driven; category-based)
  - Transcriptomics (*in vivo* short-term animal)
  - High-throughput bioactivity
- Although not NAM per se, transparency principles of systematic review and integration of toxicity pathway (e.g., AOP or MOA) information also paramount

Read-Across

Expert-driven read-across

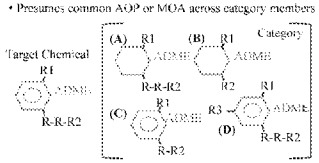
- 'Many-to-one approach'
- Approach is based on evidence across three information tiers (e.g., structural and physicochemical; toxicokinetic; and toxicity/bioactivity) to select analog(s)
- Hazard and dose-response information (e.g., point-of-departure [POD]) from single best analog used as surrogate for target chemical



⇨ POD from single best analog adopted as surrogate for data-poor target chemical

Category based read-across

- 'One-to-many' approach
- Based primarily on structural and physicochemical properties
- Robustness of approach dependent on density of analogs populating a category
- Highly reliant on weight-of-evidence supporting toxicity endpoints across category
- Presumes common AOP or MOA across category members



Tox for category
(A)= liver
(B)= kidney, liver
(C)= liver
(D)= GI, liver, kidney

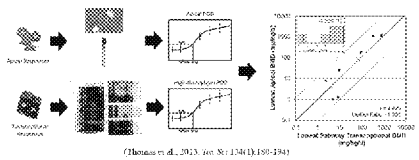
ADME = Absorption, Distribution, Metabolism, Elimination

HHRA application(s)

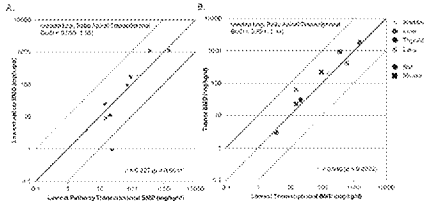
- Provisional Peer-Reviewed Toxicity Value (PPRTV) assessments; Superfund Technical Support memos to EPA Regions

Transcriptomics

Transcriptional perturbations and apical endpoints for both cancer and noncancer are evaluated in same organ tissues following short-term (e.g., 2-week) exposures

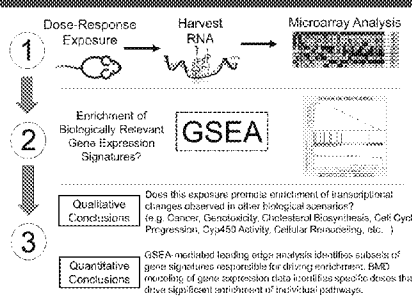


(Thomas et al., 2013, Tox Sci 134(1):180-194)



- Transcriptional pathway-based points-of-departure (PODs) from short-term *in vivo* assays were within 2-3 fold of both non-cancer (A) and cancer (B) apical PODs across different species, routes of exposure, durations of exposure, and target organ tissues
- Major challenge: relevance of transcriptional pathway perturbations to target organ toxicity?

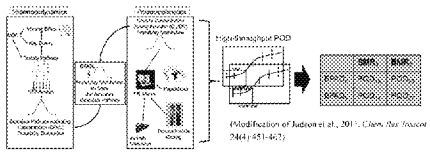
GSEA: Identifying Biologically Relevant Transcriptional Alterations



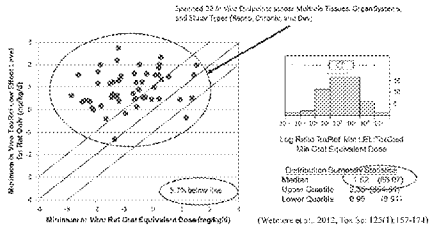
(Image courtesy of Dr. Jeffrey Dean, EPA/ORD/NCEA-Cincinnati)

High-throughput Bioactivity

Integration of in vitro biological activity data (e.g., ToxCast/Tox21) and reverse toxicokinetic in vitro to in vivo extrapolation may facilitate identification of oral equivalent doses that can be benchmark dose modeled for identification of HTP-based PODs



(Modification of Judson et al., 2011, *Chem. Res. Toxicol.* 24(4):451-462)



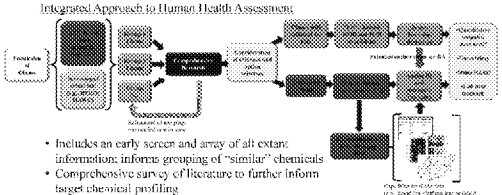
(Wetmore et al., 2012, Tox Sci 125(1):157-174)

HHRA application(s)

- Superfund Technical Support memos to EPA Regions; bioactivity information used as qualitative support for augmenting weight-of-evidence in analog(s) selection

Bringing it all together

Integrated Approach to Human Health Assessment



- Includes an early screen and array of all extant information; informs grouping of "similar" chemicals
- Comprehensive survey of literature to further inform target chemical profiling
- Traditional assessment applied when data allows
- Expert-driven read-across when hazard/dose-response data are lacking
- Integration of information from NAM data streams to fill gaps

**The collective Agency efforts presented here are in response to the NAS' suggestion to put research/processes in place to adapt to new and emerging methods

U.S. Environmental Protection Agency
Office of Research and Development

EPA United States Environmental Protection Agency

Combining data within species: Meta-analysis in IRIS

J. Allen Davis[1] and Leonid Kopylev[2]
[1]US EPA, Office of Research and Development, National Center for Environmental Assessment - Cincinnati
[2]US EPA, Office of Research and Development, National Center for Environmental Assessment - Washington

Introduction

Often, human health risk assessments have relied on qualitative approaches for hazard identification, which involves weight of evidence determinations that integrate evidence across multiple studies. In 2014, the National Research Council recommended that IRIS develop and apply quantitative approaches for evidence integration, including the application of meta-analyses to animal and human data, to help summarize and evaluate the results of a systematic review. In the meta-analytic approach, a pooled effect size is calculated after consideration of multiple potential confounding factors in order to determine whether the entire database under consideration indicates a chemical is a hazard. Two examples demonstrate approaches used in IRIS assessments: TMB (trimethylbenzene) neurotoxic hazard and pleural plaques effect on lung function.

Trimethylbenzene and pain sensitivity: methods

- A publically available, comprehensive literature search was performed in support of the IRIS Toxicological Review of trimethylbenzenes (TMBs)
- Six neurotoxicity studies were found that investigated decreased pain sensitivity following exposure in individual TMB isomers or a mixture thereof (i.e., C-9 fraction) - studies differed in testing time, test agent, and application of foot shock
- Qualitative hazard identification concluded the pain sensitivity was a hazard and that testing time mainly influenced observation of effect
- Methods outlined in Vesterinen et al. (2014) and Viechtbauer (2010) were applied using the Metafor R package
  - Random and mixed-effects models were run
  - Effect sizes were calculated as standardized mean differences
  - Hedge's G was used to account for bias due to small sample sizes
  - Restricted maximum likelihood was used to calculate total heterogeneity to prevent underestimated/biased estimates of variance
  - Publication bias, normality of residuals and sensitivity analyses were investigated

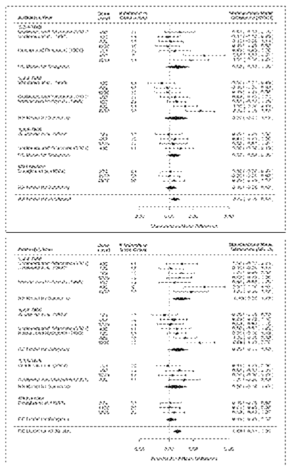


Figure 1. Forest plots for pain sensitivity studies. (A) Pre-foot shock; (B) post-foot shock

Trimethylbenzene and pain sensitivity: results

Table 1-- Meta-regression results for TMB and mixture studies included in analysis

Moderating Variable	Effect Size (95% CI) Model 1	Model 2	Number of Animals	Number of Groups	Adjusted R²	Q_E p-value	Q_M p-value
Pre-foot Shock Results							
Pooled Estimate	0.479 (0.205-0.556)*		283	22			
Lab							
Nufer	0.550 (0.258-0.841)*	Ref	223	19	0.0%	0.006	0.624
Douglas	0.259 (-0.752-1.060)	-0.291 (-1.201-0.619)	60	3			
Testing Time							
0 days	1.231 (0.914-1.548)***	Ref	69	6	95.9%	0.835	0.002
1 day	0.258 (-0.104-0.740)	-1.075 (-1.652--0.498)**	60	3			
50 days	0.506 (0.258-0.747)	-0.725 (-1.056--0.395)***	154	12			
Isomer							
1,2,4-TMB	0.647 (-0.099-1.393)	Ref	55	7	0.0%	0.003	0.845
1,2,3-TMB	0.575 (-0.121-1.265)	-0.074 (-1.093-0.944)	51	7			
1,3,5-TMB	0.396 (-0.575-1.368)	-0.261 (-1.027-0.255)	37	4			
C9 fraction	0.159 (-0.616-1.147)	-0.406 (-1.725-0.749)	60	3			
Dose	0.0001 (-0.0006-0.0005) for each 10 mg/m³ increase				0.0%	0.006	0.716
Post-foot Shock Results							
Pooled Estimate	0.663 (0.314-1.012)		283	22			
Lab							
Nufer	0.772 (0.397-1.146)***	Ref	223	19	22.02%	0.042	0.152
Douglas	0.159 (-0.619-0.936)	-0.613 (-1.476-0.250)	60	3			
Foot shock							
Yes	0.911 (0.395-1.426)**	Ref			0.00%	0.024	0.190
No	0.467 (0.004-0.929)*	-0.444 (-1.137-0.249)					
Testing Time							
0 days	1.361 (0.910-1.809)***	Ref	69	6	73.65%	0.238	0.015
1 day	0.155 (-0.467-0.754)	-1.203 (-2.065--0.271)**	60	3			
50 days	0.459 (0.094-0.844)*	-0.902 (-1.572--0.232)*	154	12			
Isomer							
1,2,4-TMB	1.004 (0.446-1.562)**	Ref	55	7	0.0%	0.030	0.420
1,2,3-TMB	0.727 (0.128-1.316)*	-0.263 (-1.168-0.594)	51	7			
1,3,5-TMB	0.551 (-0.320-1.361)	-0.459 (-1.560-0.562)	37	4			
C9 fraction	0.159 (-0.555-0.952)	-0.845 (-1.900-0.209)	60	3			
Dose	0.0001 (-0.0004-0.0005) for each 10 mg/m³ increase				0.0%		0.807

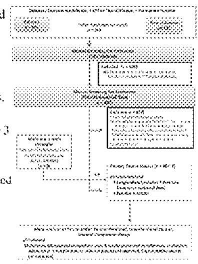
[illegible]

- Quantitative meta-analyses and meta-regressions supported original qualitative hazard identification determination -- ***decreased pain sensitivity is a hazard in humans following exposure to trimethylbenzene isomers***
- Time of testing appeared to be the study-level variable that most strongly affected differing study results and explained the majority of inter-study heterogeneity

Pleural plaques effect on lung function: methods

A literature search was conducted using the PubMed and Web of Science databases with no publication date limitations. Studies were excluded if

- the plaques group included individuals with diffuse pleural thickening (DPT)
- undefined pleural or parenchymal abnormalities.



Each paper was reviewed independently by 2 of the 3 reviewers. In cases of disagreements, the 3rd reviewer reviewed the paper and participated in the consensus- building discussions. Reviewers evaluated potential limitations in 5 aspects of study design:

- selection of participants
- protocols for x ray or HRCT readings
- protocols for spirometry measurements;
- analytic approach
- considerations of smoking.

The Metaphor R package was used for the meta-analyses

- A random effects model was used for both FVC and FEV1
- To assess possible publication bias, funnel plots were evaluated. Additional sensitivity analyses were conducted to evaluate the potential effect of identified limitations on the meta-analyses results.

Pleural plaques effects on lung function: results

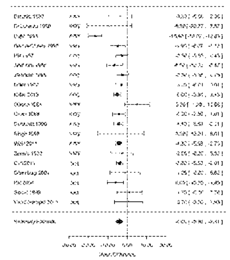


Figure 1 (FVC)

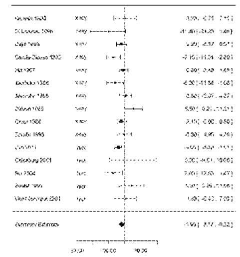


Figure 2 (FEV1)

- The summary effect estimates for both FVC and FEV1 are statistically significant, showing a change of −4.09 %pred (95% CI: −5.86, −2.31) and −1.99 %pred (95% CI: −3.77, −0.22), respectively (See Fig. 1 and Fig. 2)
- The results of larger studies are very consistent in showing a decrease in FVC (see Fig. 1). In contrast, fewer large studies are available for FEV1, and there is less consistency in the results (see Fig. 2).
- At the individual level, the decrement in FVC or FEV1 may or may not be noticeable for a given patient; while many with pleural plaques could have well-preserved lung function, there are some at the lower end of 'normal' lung function, for whom even a small additional decrement would result in an increased in disease severity (e.g., mild to moderate disease).
- At the population level, even small changes in the average of a distribution of lung function can result in a proportion of the exposed population shifted down into the lower "tail" of the distribution, into clinically significant lung function deficit region

Discussion

- Both human and animal data are amenable to quantitative synthesis via meta-analysis
- Studies need not be exactly the same, as long as results are reported in a consistent way or can be converted into a comparable format (e.g., use of standardized mean difference as effect metric)
- Use of free R software allows conducting meta-analysis
- Use of meta-analytic methods for hazard identification are in line with National Research Council (2014) recommendations for the development of quantitative hazard identification and evidence integration methods
- Applying meta-analysis and meta-regression methods will help to improve future risk assessments and ensure the use of the best available science

References:

Davis JA, Kraft A. Quantitative meta-analytic approaches for the systematic synthesis of data and hazard identification: a case study of decreased pain sensitivity due to trimethylbenzene exposure. 2017. *Environmental Research*. 158: 598-609.

Kopylev L, Christensen KY, Brown JS, Cooper GS. A systematic review of the association between pleural plaques and changes in lung function. 2015. *Occupational and Environmental Medicine*. 72(8): 606-14.

EPA United States Environmental Protection Agency

Systematic Evaluations of Physiologically-Based Pharmacokinetic Models for Human Health Risk Assessment

Alan F. Sasso[1], Paul M. Schlosser[1]
[1]U.S. EPA, Office of Research and Development, National Center for Environmental Assessment

Background

- Physiologically-based pharmacokinetic (PBPK) models are tools for estimating absorption, distribution, metabolism, and elimination (ADME) of chemicals in the body
  - Quantify internal (tissue/organ) dose vs exposure
  - Facilitate dose-response analysis/human extrapolation
  - Use chemical- and species-specific data (unlike default $BW^{3/4}$ allometric scaling)
- Multiple alternative models or analyses in literature
  - "Being published is not enough": EPA thoroughly evaluates models based on scientific and technical criteria prior to use in an assessment
  - IRIS uses a structures approach to evaluate quality and usability
- The evaluation process stresses: (1) clarity in the documentation of model purpose, structure, and biological characterization; (2) validation of mathematical descriptions, parameter values, and computer implementation; and (3) evaluation of each plausible dose metric.
- **NAS (2014) recommendations addressed**
  - **Develop and expand use of formal quantitative methods in data integration for dose-response assessment and derivation of toxicity values**
  - **Develop tools for assessing risk of bias in different types of studies**

Identification and Inventory of PBPK Models

- A thorough literature search is conducted to identify existing PBPK models
- A summary report is prepared of available models and their possible utility for use **(scoping)**
  - This work is conducted by the **Pharmacokinetics Workgroup (PKWG)***, in conjunction with information specialists
- Table 1 outlines typical summary information presented for each model at the scoping phase

Table 1. Example animal PBPK inventory table for model scoping

Author	Smith et al. (2003)				
Contact Email	xxxxx@email.com				
Contact Phone	xxx-xxx-xxxx				
Sponsor	N/A				
Model Summary					
Species	Rat				
Strain	F344				
Sex	Male and female				
Life-Stage	Adult				
Exposure Routes	Inhalation	Oral	I.V.	Skin	
Tissue Dosimetry	Blood	Liver	Kidney	Urine	Lung
Model Evaluation					
Language	ACSL 11.6				
Code Available	YES	Effort to Recreate Model		COMPLETE	
Code Received	YES	Effort to Migrate Code		SIGNIFICANT	
Structure Evaluated	YES				
Math Evaluated	YES				
Code Evaluated	YES. Issue (minor): Incorrect units listed in comments for liver metabolism (line 233). Issue (major): Mass balance error in stomach compartment				
Available PK Data	Urine (cumulative amount excreted) and blood (concentration) time course data for oral (gavage) and inhalation (3hr/day for 4 days) exposure. In vitro skin permeation.				

*The Pharmacokinetics Workgroup (PKWG) is convened by the National Center for Environmental Assessment (NCEA) to support and promote consistent application of the best science in the field of mathematical modeling of pharmacokinetic processes and the data supporting it as applied in human health risk assessment. It is composed of scientists with specific expertise in the range of disciplines involved in the construction and development of pharmacokinetic models, evaluation of data supporting such models, statistical analysis of data and modeling results, and characterization of related uncertainty and variability.

Evaluation of PBPK Models

PBPK Model Scoping: Criteria A

- An evaluation of a model is required before accepting it for use in an assessment
  - Many models contain errors with varying degrees of impact on model predictions
- Initial judgments on the suitability of a model are separated into two categories: scientific and technical (Table 2)

Table 2. Evaluation criteria for PBPK models

Criteria	Example information
Scientific	Biological basis for the model is accurate. Consistent with mechanisms that significantly impact dosimetry. Predicts dose metrics expected to be relevant. Applicable for relevant route(s) of exposure.
	Consideration of model fidelity to the biological system strengthens the scientific basis of the assessment relative to standard exposure-based extrapolation (default) approaches. Can the model describe critical behavior, such as nonlinear kinetics in a relevant dose range, better than the default (i.e., $BW^{3/4}$ scaling)? Is the available metric a better predictor of risk than default? Specifically, model-based metrics may correlate better than the applied doses with animal/human dose-response data. Degree of certainty in model predictions vs. default is also a factor. For example, while target tissue metrics are generally considered better than blood concentration metrics, lack of data to validate tissue predictions when blood data are available may lead to a choice of the latter.
	Principle of parsimony. Model complexity or biological scale, including number and parameterization of (sub)compartments (e.g., tissue or subcellular levels) should be commensurate with data available to identify parameters.
	Model describes existing PK data reasonably well, both in "shape" (matches curvature, inflection points, peak concentration time, etc.) and quantitatively (e.g., within a factor of 2–3).
	Model equations are consistent with biochemical understanding and biological plausibility.
Initial technical	Well-documented model code is readily available to EPA and public.
	Set of published parameters clearly identified, including origin/derivation.
	Parameters do not vary unpredictably with dose (e.g., any dose dependence in absorption constants is predictable across the dose ranges relevant for animal and human modeling).
	Sensitivity and uncertainty analysis has been conducted for relevant exposure levels (local sensitivity analysis is sufficient, though global provides more information). If a sensitivity analysis was not conducted, the PKWG would suggest this as additional work before using the model in the risk assessment. A sound explanation should be provided when sensitivity of the dose metric to model parameters differs from what is reasonably expected based on experience.
$BW^{3/4}$ = body-weight scaling to the 3/4 power; PK = pharmacokinetic; PKWG = Pharmacokinetic Working Group.	

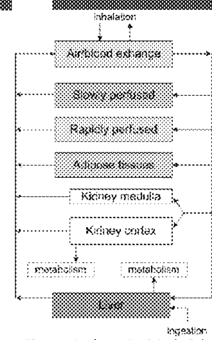
In-Depth Technical Evaluation: Criteria B

- Primarily address computational implementation and technical issues
- Only conducted on models that pass review for Criteria A
  - Criteria B evaluation is not possible without model code
- Model equations and parameters in computer codes match those in the manuscript or report
- Published figures/tables of model simulations are reproducible using the available code (within 10% of the publication).
- If errors in model code or parameters are found and corrected, the revised model must still be in agreement with data. Errors must be small enough to not invalidate the model, parameters, or assumptions.
  - Model predictions outside the range of the data are allowed to change by more than 10% of the original model or publication, since this would be considered a model correction.

Resource Considerations for PBPK Model Revision or Development: Criteria C

If existing models fail Criteria A or B, the potential value in implementing a PBPK in a risk assessment must be weighed against the time, effort, and possible expenses required to address model shortcomings.

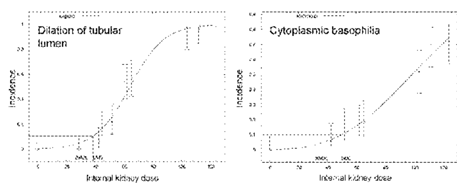
PBPK Evaluation Example: Chloroform



- Chloroform is a trihalomethane present in drinking water as a byproduct of disinfection
- Kidney and liver are target organs
  - Effects induced via production of reactive metabolites
- A PBPK model (left) was obtained during scoping and satisfied Criteria A
- Issues identified during in-depth technical evaluation:
  - Metabolic parameter derivation for the kidney cortex contained a units error, and the calculation was not performed consistently for humans and rodents
  - The volume ratio for kidney cortex and medulla was reversed in the code, and did not match the reported value or original reference

Upon evaluation under Criteria C, it was determined:

- Time and effort to correct the model was minimal
- Corrections led to little or no changes in model predictions of data
  - Estimates of the internal dose metric (kidney metabolism) changed significantly. Since there are no *in vivo* data available for this measure, this was considered a correction to the original model.
- Model was successfully revised by EPA, published as journal article (Sasso et al., 2013)



The revised PBPK model allows for improved quantitative dose-response modeling and data integration. Kidney endpoints can be evaluated across different routes of exposure and different species (Nagano et al., 2006, and Yamamoto et al., 2002). The figures above illustrate dose-responses for rats from multiple exposure routes (inhalation, oral, and combined inhalation+oral) on basis of PBPK-derived kidney dose.

Selected references

McLanahan et al. (2012). Physiologically based pharmacokinetic model use in risk assessment--Why being published is not enough. *Tox. Sci.*, 126: 5-15.

Nagano, et al. (2006). Enhancement of renal carcinogenicity by combined inhalation and oral exposures to chloroform in male rats. *J. Toxicol. Environ. Health Part A* 69, 1827–1842.

Sasso et al. (2013). Application of an updated physiologically based pharmacokinetic model for chloroform to evaluate CYP2E1-mediated renal toxicity in rats and mice. *Tox. Sci.*, 131: 360-374.

Yamamoto, et al. (2002). Carcinogenicity and chronic toxicity in rats and mice exposed to chloroform by inhalation. *J. Occup. Health* 44, 283-293.

U.S. Environmental Protection Agency
Office of Research and Development

EPA United States Environmental Protection Agency

Male reproductive toxicity in animal studies of diisobutyl phthalate (DIBP): a case study application of systematic review approaches

Erin E. Yost[1], Susan Y. Euling[1], James A. Weaver[1], Brandiese E. J. Beverly[1], Nagalakshmi Keshava[1], Anuradha Mudipalli[1], Xabier Arzuaga[2], Todd Blessinger[2], Laura Dishaw[1], Andrew Hotchkiss[1], Susan L. Makris[2]

[1] US EPA National Center for Environmental Assessment, Research Triangle Park, NC; [2] US EPA National Center for Environmental Assessment, Washington, DC

Introduction

Diisobutyl phthalate (DIBP) is used as a plasticizer in a variety of industrial and consumer products. Although DIBP has been less widely studied compared to other phthalates, there is evidence that DIBP and its primary metabolite, monoisobutyl phthalate (MIBP), cause male reproductive toxicity. A recent systematic review of endocrine-related low-dose toxicity by the National Academies of Sciences (NAS) evaluated the effects of DIBP on three anti-androgenic outcomes [testosterone, anogenital distance (AGD), and hypospadias], and concluded that DIBP is a presumed human hazard based on decreased fetal testosterone in rodents exposed during gestation. The Integrated Risk Information System (IRIS) performed a systematic review of male reproductive effects of DIBP exposure that considered all outcomes and all life stages of exposure, following recommendations in the 2014 NAS review of the IRIS program. Here, we use studies that evaluated testosterone in male rodents exposed to DIBP or MIBP as a case study of the IRIS systematic review process. We also summarize the overall conclusions for male reproductive effects identified in the IRIS systematic review of DIBP, and compare these results to the findings of NAS.

Methods

Animal studies for DIBP or MIBP were identified by searching four online databases (PubMed, Web of Science, Toxline, and TSCATS2), using search terms designed to capture all potentially pertinent studies. The last update was in July 2017. Title/abstract screening was used to identify primary health effect studies that exposed non-human mammalian animals to any administered dose of DIBP or MIBP via oral, dermal, or inhalation routes. These studies were evaluated by at least two reviewers using the approach in Figure 1.

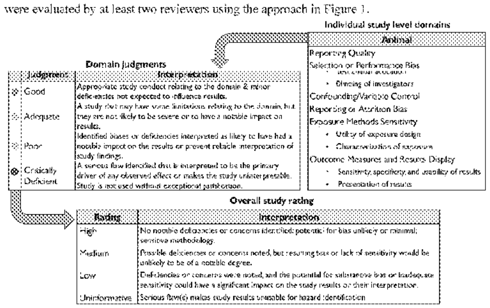


Figure 1. Study evaluation process

After study evaluation, the evidence for each health effect outcome was synthesized according to the developmental stage of exposure. Based on this synthesis, the evidence was assigned a conclusion of *robust, moderate, slight, indeterminate,* or *compelling evidence of no effect.* The ratings for individual outcomes were summarized into an overall conclusion for male reproductive effects using a structured framework (see Poster by Yost et al.).

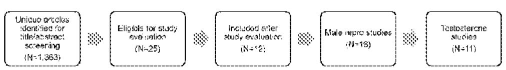


Figure 2. Abbreviated literature flow diagram

U.S. Environmental Protection Agency
Office of Research and Development

Results

Table 1. Animal studies of testosterone and DIBP or MIBP exposure. Of the 11 studies that evaluated testosterone in male rats or mice, 7 exposed animals during gestation and/or until weaning, and 4 were postnatal exposures of males near the time of puberty. The postnatal exposure studies had higher risk of bias because of reporting limitations, including uncertainty about the pubertal status of the test animals at the time of exposure.

Reference	Study description			Study evaluation									
	Population	Exposure	Outcome	Reporting quality	Test animal allocation	Blinding of investigators	Confounding / variable control	Reporting or attrition bias	Characterization of exposure	Utility of exposure design	Sensitivity, specificity, and usability of results	Presentation of results	Overall confidence
Borch et al. 2006	Rat (Wistar)	Diet GD 7-19	Fetal T prod/conc	[illegible]	[illegible]	A	[illegible]	[illegible]	A	A	[illegible]	[illegible]	[illegible]
Howdeshell et al. 2008	Rat (Sprague-Dawley)	Gavage GD 8-18	Fetal T prod	[illegible]	A	A	[illegible]	[illegible]	A	[illegible]	[illegible]	[illegible]	[illegible]
Saillenfait et al. 2017	Rat (Sprague-Dawley)	Gavage GD 13-19	Fetal T prod	[illegible]	[illegible]	A	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Furr et al. 2014	Rat (Sprague-Dawley)	Gavage GD 14-18	Fetal T prod	[illegible]	[illegible]	A	[illegible]	A	A	[illegible]	[illegible]	[illegible]	[illegible]
Hannas et al. 2012	Rat (Sprague-Dawley)	Gavage GD 14-18	Fetal T prod	[illegible]	A	A	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	A	[illegible]
Hannas et al. 2011	Rat (Sprague Dawley)	Gavage GD 14-18	Fetal T prod	[illegible]	A	A	[illegible]	[illegible]	A	[illegible]	[illegible]	[illegible]	[illegible]
Wang et al. 2017	Mouse (ICR)	Diet GD 0-21; GD 0-PND 21	Postnatal and Adult T conc	[illegible]	[illegible]	[illegible]	A	A	A	[illegible]	A	P	Medium
Oishi and Hiraga 1980a	Mouse (JCL:ICR)	Diet PND 35-42	Postnatal T conc	A	NR	NR	A	[illegible]	A	P	A	A	Medium
Oishi and Hiraga 1980b	Mouse (JCL:ICR)	Diet PND 35-42	Postnatal T conc	A	NR	NR	A	P	A	P	A	A	Medium
Hiraga 1980c	(JCL:Wistar)	PND 35-42	T conc										
Oishi and Hiraga 1980d	Rat (JCL:Wistar)	Diet PND 35-42	Postnatal T conc	A	NR	NR	A	P	A	A	A	A	Medium

Abbreviations: Gestation day (GD), Postnatal day (PND), Testosterone (T) production (prod) or concentration (conc)

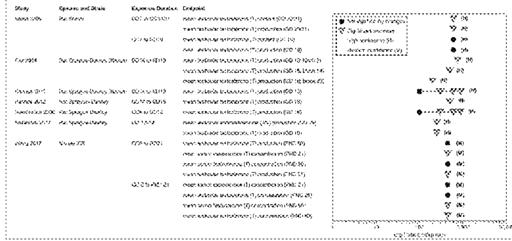


Figure 3. Summary of exposure-response for testosterone from gestational exposure studies.

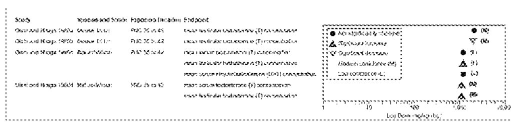


Figure 4. Summary of exposure-response for testosterone from postnatal exposure studies.

The synthesis of results for testosterone is summarized in an evidence profile table (Table 2). Gestational exposure studies provided *robust* evidence for effects on testosterone, whereas evidence from postnatal exposure studies was found to be *indeterminate*. Evidence judgments for other male reproductive endpoints identified in this systematic review are summarized in Table 3.

Table 2. Evidence profile table for animal studies of testosterone and DIBP

	Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings
Gestational exposure	**High confidence** Borch et al. 2006; Furr et al. 2014; Hannas et al. 2011; Hannas et al. 2012; Howdeshell et al. 2008; Saillenfait et al. 2017; **Medium confidence** Wang et al. 2017	• Consistency • Exposure-response gradient • Effect size • Biological plausibility (support from mechanistic evidence) • Minimal concern for bias		⊕⊕⊕ ROBUST A dose-related decrease in testicular androgen levels or production (up to -96% compared to control) was observed in all studies in rats and mice that evaluated this endpoint. Several of these studies also demonstrate decreased testicular expression of genes and proteins in the steroidogenesis pathway, which provides support for biological plausibility.
Postnatal exposure	**Medium confidence** Oishi and Hiraga 1980a; Oishi and Hiraga 1980b; Oishi and Hiraga 1980c; **Low confidence** Oishi and Hiraga 1980c	• Biological plausibility	• High risk of bias • Unexplained inconsistency	○○○ INDETERMINATE A dose related increase in androgen levels was observed in two rat studies (Oishi and Hiraga 1980c-d), whereas androgen levels were decreased or not changed in mice (Oishi and Hiraga 1980a-b).

[illegible] reproductive toxicity following DIBP exposure

Outcome	Includes these endpoints	Evidence following gestational exposure	Evidence following postnatal exposure
Testosterone	Androgen levels	Robust	Indeterminate
Male morphological development	AGD, nipple retention, preputial separation, hypospadias, cleft prepuce, exposed os penis, cryptorchidism	Robust	N/A
Sperm evaluation and histopathological effects in testis or epididymis	Sperm concentration and motility, oligospermia, azoospermia, granulomatous inflammation, tubular degeneration, tubular necrosis, interstitial hyperplasia	Robust	Moderate
Reproductive organ weight	Testis, epididymis, seminal vesicle weights	Moderate	Moderate
Male reproductive overall		Robust	

Discussion

Overall, the results from animal studies of male reproductive effects provide robust evidence of a hazard from DIBP exposure. Conclusions for testosterone are consistent with those of NAS (2017). The NAS review was limited to gestational exposure studies and excluded studies that exposed animals to a single high dose (≥500 mg/kg-day); therefore, NAS only considered two fetal testosterone studies, and had inadequate evidence to evaluate the effects of DIBP on AGD or hypospadias. The IRIS systematic review included all dose levels and life stages of exposure, and was able to evaluate a wider range of androgen-dependent and -independent male reproductive outcomes. *Disclaimer: The views expressed in this poster are those of the author(s) and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.*

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.

EPA United States Environmental Protection Agency

Male reproductive toxicity in epidemiology studies of phthalates: a case study application of systematic review approaches

Elizabeth Radke, Glinda Cooper

U.S. EPA, Office of Research and Development (ORD), National Center for Environmental Assessment (NCEA), IRIS Division

Introduction

Phthalates have anti-androgenic activity in rodents resulting in reduced circulating testosterone and male reproductive tract abnormalities. Several epidemiologic studies have examined this association in humans. The National Academies of Sciences (NAS) recently published a systematic review of endocrine-related low-dose toxicity that included examination of phthalates and male reproductive tract development, and the Integrated Risk Information System (IRIS) performed a systematic review of all male reproductive effects of phthalate exposure, following recommendations in the 2014 NAS review of the IRIS program. Here, we use the associations between anogenital distance (AGD) in humans and two phthalates, di(2-ethylhexl phthalate (DEHP) and diisobutyl phthalate (DIBP), as a case study of the IRIS systematic review process. We also compare our conclusions to those of the NAS and summarize our overall findings on epidemiology studies of male reproductive effects of phthalates.

Methods

Epidemiology studies were identified by conducting a single broad literature search on the six phthalates of interest. The following databases were searched: PubMed, Web of Science, and Toxline. The last update was in January 2017. Title/abstract and full text screening was performed by two reviewers. Studies were evaluated by at least two reviewers using the approach in Figure 1.

Individual study level domains

Epidemiology
Exposure measurement
Outcome ascertainment
Population Selection
Confounding
Analysis
Sensitivity
Selective reporting

Domain judgments

Judgment	Interpretation
Good	Appropriate study conduct relating to the domain & minor deficiencies not expected to influence results.
Adequate	A study that may have some limitations relating to the domain, but they are not likely to be severe or to have a notable impact on results.
Poor	Identified biases or deficiencies interpreted as likely to have had a notable impact on the results or prevent reliable interpretation of study findings.
Critically Deficient	A serious flaw identified that is interpreted to be the primary driver of any observed effect or makes the study uninterpretable. Study is not used without exceptional justification.

Overall study rating

Rating	Interpretation
High	No notable deficiencies or concerns identified; potential for bias unlikely or minimal; sensitive methodology.
Medium	Possible deficiencies or concerns noted, but resulting bias or lack of sensitivity would be unlikely to be of a notable degree.
Low	Deficiencies or concerns were noted, and the potential for substantive bias or inadequate sensitivity could have a significant impact on the study results or their interpretation.
Uninformative	Serious flaw(s) makes study results unusable for hazard identification

Figure 1. Study evaluation process

After study evaluation, the evidence for each outcome was synthesized for each phthalate, considering aspects of an association that may suggest causation. Based on this, the evidence was assigned within stream confidence judgments of *robust, moderate, slight, indeterminate,* or *compelling evidence of no effect*. The judgments for individual outcomes were summarized into an overall conclusion for male reproductive effects using a structured framework (see Poster by Yost et al.).

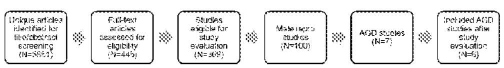


Figure 2. Abbreviated literature flow diagram

U.S. Environmental Protection Agency
Office of Research and Development

Results

Table 1. Epidemiology studies of AGD and phthalate exposure

	Reference	Population	Exposure	Outcome	Exposure	Outcome	Selection	Confounding	Analysis	Overall confidence
		Study description			Study evaluation					
Included	Bornehag et al., 2015	Birth cohort (N=196 boys) in Sweden	Single urine sample (1st trimester)	AGD at 19-21 mo	A/P	G	G	A	G	Medium
Included	Bustamante-Montes et al., 2013	Birth cohort (N=73 boys) in Mexico	Single urine sample (3rd trimester)	AGD at 1-2 d	P	[illegible]	A	A	G	Low
Included	Jensen et al., 2016	Birth cohort (N=273 boys) in Denmark	Single urine sample (28-30 wk gestation)	AGD at 3 mo	A/P	G	G	A	G	Medium
Included	Suzuki et al., 2012	Birth cohort (N=73 boys) in Japan	Single urine sample (3rd trimester)	AGD at 1-3 d	P	A	P	P	A	Low
Included	Swan, 2008	Birth cohort (N=106 boys) in U.S.	Single urine sample (mean 28 wk gestation)	AGD at 1-36 mo	A/P	P	[illegible]	A	P	Low
Included	Swan et al., 2015	Birth cohort (N=365 boys) in U.S.	Single urine sample(1st trimester)	AGD at 1-2 d	A/P	[illegible]	A	A	A	Medium

G=good; A=adequate, P=poor, A/P=adequate for short chain phthalates, poor for long chain.
Studies with biomarker measures based on samples other than urine (e.g., blood) were considered to be critically deficient for all short-chain phthalates and for primary metabolites (e.g., MEHP, MiNP) of long-chain phthalates.

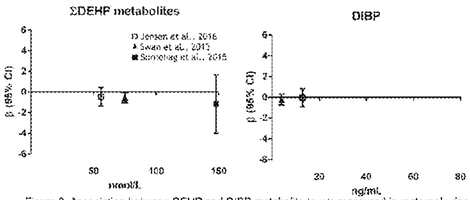


Figure 3. Association between DEHP and DIBP metabolite levels measured in maternal urine samples during pregnancy and AGD in boys in medium confidence studies

Regression coefficients on the y-axis are plotted against exposure level on the x-axis (population median for each study).

Table 2. Evidence profile table for epidemiology studies of AGD and DEHP and DIBP

	Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings and within stream evidence judgment
DEHP	**Medium confidence** Bornehag et al., 2015 Jensen et al., 2016 Swan et al., 2015 **Low confidence** Bustamante-Montes et al., 2013 Suzuki et al., 2012 Swan, 2008	Among medium confidence studies: • consistency • exposure-response gradient across studies • minimal concerns for bias	• low precision in study with largest effect size	⊕⊕◯ MODERATE Inverse associations between DEHP exposure and anogenital distance reported in 5/6 studies (Jensen et al., 2016, Swan et al., 2015, Bornehag et al., 2015, Swan, 2008, Suzuki et al., 2012), of which 2 were statistically significant (Swan et al., 2015, Swan, 2008). Among the 3 medium confidence studies, effect size increased with increasing exposure levels.
DIBP	**Medium confidence** Jensen et al., 2016 Swan et al., 2015 **Low confidence** Swan, 2008	• low study sensitivity may explain lack of association	• inconsistency • few studies	⊕◯◯ SLIGHT Inverse associations between DIBP exposure and anogenital distance reported in 2/3 studies (Swan, 2008, Swan et al., 2015), though neither were statistically significant. Exposure levels and range were low in all studies.

Of the seven identified studies on phthalates and AGD (Figure 2), one was excluded due to inadequate exposure measurement. Summary of the evaluations for the six included studies is in Table 1. Results of medium confidence studies were given priority (Figure 3), but all studies were included in the synthesis, which is summarized in the evidence profile table (Table 2). For DEHP, an exposure response gradient was observed across studies, with the study with the highest exposure levels reporting the strongest association. This was not observed for DIBP, but exposure levels were low in all studies. The same methods were used for other phthalate/outcome combinations and the within stream evidence judgments are shown in Figure 4. Table 3 presents a comparison of the within stream judgments from the IRIS and NAS reviews of anogenital distance, testosterone in infants, and hypospadias. Both found that the evidence for the latter two outcomes was not adequate to form a conclusion. For anogenital distance, evidence for DEHP and DBP was considered *moderate* in both reviews. Evidence for DINP, DIBP, and BBP was considered *slight* by IRIS and *inadequate* by NAS. These conclusions were not considered inconsistent, but rather reflect differences in the process for evidence synthesis. Only DEP differed between reviews, classified as *slight* by IRIS and *moderate* by NAS based on the results of a meta-analysis.

Outcome	DEHP	DINP	DBP	DIBP	BBP	DEP
Anogenital distance	M	S	M	S	S	S
Hypospadias/cryptorchidism	I	S	S	S	S	I
Pubertal development	S	S	S	S	S	S
Semen parameters	M	M	[illegible]	S	M	S
Time to pregnancy	S	I	M	S	M	I
Testosterone	M	M	S	M	I	I
Male repro overall	[illegible]	M	[illegible]	M	M	S

Robust (R) Moderate (M) Slight (S) Indeterminate (I)

Figure 4. Within stream evidence judgments for human evidence of male reproductive effects associated with phthalates

Table 3. Within stream evidence judgments of systematic reviews of male reproductive developmental toxicity in epidemiology studies by IRIS and NAS

Phthalate	Anogenital distance IRIS	Anogenital distance NAS	Testosterone in infants IRIS	Testosterone in infants NAS	Hypospadias IRIS	Hypospadias NAS
DEHP	Moderate	Moderate	Indeterminate	Inadequate	Indeterminate	Inadequate
DINP	Slight	Inadequate	Indeterminate	Inadequate	Indeterminate	Inadequate
DBP	Moderate	Moderate	Indeterminate	Inadequate	Indeterminate	Inadequate
DIBP	Slight	Inadequate	Indeterminate	Inadequate	Indeterminate	Inadequate
BBP	Slight	Inadequate	Indeterminate	Inadequate	Indeterminate	Inadequate
DEP	Slight	Moderate	Indeterminate	Inadequate	Indeterminate	Inadequate

Classifying levels: IRIS: Robust, Moderate, Slight, or Indeterminate; NAS: High, Moderate, Low, or Inadequate

Discussion

Overall, the results from epidemiology studies of male reproductive effects provide evidence of a hazard from phthalate exposure. Looking specifically at anogenital distance, there is *moderate* evidence of an association with DEHP and DBP exposure, and *slight* evidence for other phthalates. These findings are generally consistent with the NAS report on low-dose toxicity from endocrine active chemicals (2017). In the case of DIBP, the weaker evidence may be largely explained by the smaller number of studies and low exposure levels that decreased study sensitivity.

Quantitative Evaluation of Uncertainty: APROBA and Beyond

Todd Blessinger and David Bussard
US EPA, Office of Research and Development, National Center for Environmental Assessment - Washington

Todd Blessinger | Blessinger.Todd@epa.gov | 202-564-1641

Purpose and Scope

Quantitative assessment of uncertainty was recommended by the NRC:

- *Science and Decisions* report (NRC, 2009) – recommended incorporating probabilistic methods for assessing uncertainty.
- *Review of the IRIS Program* report (NRC, 2014) – recommended systematic use of uncertainty analysis and expanded use of Bayesian methods.

NCEA will pilot this approach to better understand issues in implementing it and to engage in dialogue with stakeholders as to advantages and challenges in utilizing this approach.

Probabilistic Calculation of Risk-Specific Doses

Goal: Probabilistically incorporate adjustments and uncertainty when extrapolating dose-response results from animal data to the human population.

Current Practice: Reference values (RfVs) are generally calculated by dividing a point of departure (POD; usually a BMDL or NOAEL) by a series of uncertainty factors (UFs):

$$\text{RfV} = \frac{\text{POD}}{\text{UF}_1 \times \cdots \times \text{UF}_k}$$

- Default values of UFs are (1, 3, or 10).
- Decision on which value to use is made qualitatively based on information available for the particular assessment (e.g., size of database, study characteristics)
- Reference Value definition does not explicitly target incidence, effect size, or confidence.

Proposed New Practice: Calculate risk-specific dose intervals using probabilistically-defined versions of POD and UFs, using the concept of target human dose.

Target Human Dose and APROBA

Target human dose, HD_M^I:

- HD_M^I = the Human Dose at which a fraction (or incidence) I of the population shows an effect of magnitude (or severity) M or greater for the critical effect considered.
  - A "risk-specific dose."

Examples:

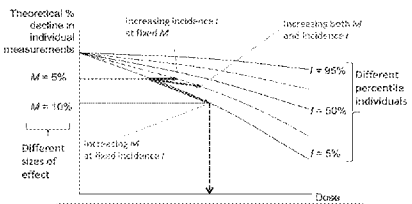
- $HD_{10}^{5\%}$ = human dose at which 1% of the population shows an increase in liver weight of 10% or greater above background.
- $HD_{50}^{5\%}$ = human dose at which there is an individual extra risk of lung tumors of 5% (or more) in 1% of the population.

HD_M^I is calculated using the formula similar to RfV:

$$HD_M^I = \frac{\text{POD}}{\text{AF}_1 \times \cdots \times \text{AF}_k} \quad (1)$$

- Each AF, or "assessment factor," is treated as a continuous random variable; the parameters of the distributions of these random variables can be determined from empirical data. The resulting HD_M^I is a random variable with its own probability distribution.

Target Human Dose (cont'd)



Key Concept: HD_M^I (e.g., $HD_{10}^{5\%}$)

Approximate Probability Analysis (APROBA) is an Excel-based tool to calculate a probabilistic RfV from animal data.

- Computes HD_M^I under the assumption that the POD and AFs are independent lognormally distributed.
- An analogue to a reference value can be derived for a pre-selected percentile (e.g., 5th percentile) of the HD_M^I distribution. The interval reflects uncertainty as well as a choice of a desired confidence (e.g., 95%) in the HD_M^I estimate.
- Was applied by the Dutch National Institute for Public Health and the Environment (RIVM) in recent risk assessment on melamine.

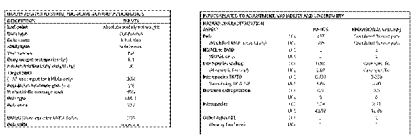
Example

Dose-response data of absolute epididymis weight in adult rats after exposure to chemical X by inhalation:

Exposure (ppm)	No. of animals	Mean (mg)	SD (mg)
0	25	0.3327	0.03631
100	25	0.3311	0.04453
250	25	0.3053	0.04188
500	25	0.2912	0.05206
750	25	0.2405	0.04804

Exponential model 3 fit to data at BMR of 10% relative deviation from control mean yields: BMDL = 237 ppm; BMDU = 535 ppm

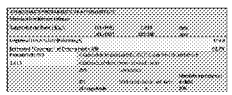
Input in APROBA worksheet:



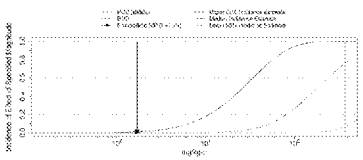
- Input on left entered by user
- Values on right are lower and upper confidence limits representing the estimated 5th and 95th percentiles of the lognormal distribution for the AFs.
  - LCL and UCL calculated using empirical data
- HDMI has lognormal distribution based on formula in Equation (1).

Example (cont'd)

APROBA output:



RfV = 1.6 ppm, which is the LCL (P05 = 5th percentile) of the HDMI distribution.



Plot: CDFs of Lower, Median, and Upper Incidence Estimates

- Several types of "central" estimates can be derived, such as the median or the expected value, if assuming a log-normal distribution.
- The approach could also be modified to provide a distribution on the population risk at a given dose.
  - Distribution can be used to estimate benefits of reduced exposures or for communication about risks of exposure.

Next Steps

- Conduct a case study using APROBA to evaluate the advantages of incorporating quantitative uncertainty in assessments with this approach.
  - Evaluate the information and choices needed to produce the estimates.
  - Work with risk managers to evaluate if this approach is useful, and how it might need modification to be more useful.
- Apply uncertainty analysis to risk assessment done to support benefit-cost analysis.
- Non-APROBA-based uncertainty analysis.

References

- IPCS (International Programme on Chemical Safety). (2014). Guidance document on evaluating and expressing uncertainty in hazard characterization. World Health Organization.
- Chiu, WA; Slob, W. (2015). A unified probabilistic framework for dose–response assessment of human health effects. Environ Health Perspect; (123) pp 1241-1254.
- Risk assessment and derivation of a provisional guideline value for melamine in drinking water. Advice to: Ministry of Infrastructure and Environment (Inspectorate of Environment and Transport) RIVM.

Printed on 100% recycled paper with a minimum 50% post-consumer fiber using vegetable based ink.

U.S. Environmental Protection Agency
Office of Research and Development

The views expressed in this poster are those of the author(s) and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency.

Mode of action and human relevance evaluation of Dibutyl Phthalate (DBP)-induced male reproductive system toxicity.

EPA United States Environmental Protection Agency

Xabier Arzuaga[1], Tenelle Walker[1], Andrew Hotchkiss[1]
[1]US EPA, Office of Research and Development, National Center for Environmental Assessment

Introduction

Dibutyl phthalate (DBP) is used as a plasticizer in a variety of commercial and consumer products (US EPA, 2014; Kavlock et al. 2002). The largest source of DBP exposure in humans is food, with inhalation and dermal exposures considered minimal (Kavlock et al 2002). Epidemiological studies provide evidence of human exposure and altered androgen levels during lifestages at which androgen production is critical for the normal development and function of the male reproductive system (WHO/UNEP, 2013), and experimental studies using rat models have reported that exposure to DBP is associated with adverse responses in the male reproductive system. Effects include decreased androgen production, agenesis of the male reproductive system and increased incidence of internal and external malformations after developmental exposures (e.g. degeneration of seminiferous tubules, hypospadias), and decreased fertility and sperm counts (CPSC, 2010; Makris et al 2013; US EPA, 2009). Evidence from post-natal exposure studies also suggests that young animals are more sensitive to phthalate-induced testicular injury than adults (Boekelheide et al 2004). However, recent studies using ex-vivo human tissue culture preparations, or rodent and human testicular tissue xenografts report that human fetal testes are resistant to phthalate induced disruption of testosterone production (Johnson et al., 2012; Albert and Jégou, 2014). Such findings raise questions about the human relevance of the androgen-related endpoints measured in experimental rodents exposed to phthalates.

A mode of action framework was used to evaluate the available evidence from experimental and in-vitro studies according to lifestage of exposure. Studies considered for this analysis include:

- Exposures during the masculinization programming window (MPW; gestational period during which development of the male reproductive system occurs).
- Exposures during early post-natal stages.

Methods

The experimental and mechanistic studies considered in this analysis were obtained from the literature search performed by the US EPA Integrated Risk Information System (IRIS). Studies for DBP or MBP were identified from online databases (PubMed, Web of Science, Toxline, and TSCATS2) using search terms designed to capture pertinent studies. The last update was performed in July 2017. Title/abstract screening followed by a full text review was performed to identify relevant studies on male reproductive effects and related mechanisms/pathways (See Figure 1 below). The types of in-vivo and in-vitro studies considered most informative to our evaluation were:

- Gestational DBP exposure studies that use mammalian in-vivo and in-vitro models, and human xenograft and ex vivo models treated during the masculinization programming window.
- Additional ex-vivo studies that expose human fetal testis tissue cultures to DEHP or its metabolite MEHP.
- Studies aimed at characterizing the receptor for DBP at a molecular level.
- Post-natal DBP exposure studies that use mammalian model species, including in-vivo, xenograft, and cell culture models.

The available mechanistic and toxicological evidence was analyzed in concordance with the framework and levels of biological organization used for mode of action Action analysis for non-cancer effects and development of Adverse Outcome Pathways (Bobbis et al 2008; Edwards et al 2016). As recommended by US EPA's Framework for Assessing Health Risk of Environmental Exposures to Children and the World Health Organization International Programme on Chemical Safety, the available mechanistic and toxicological studies and endpoints that inform the mode of action for DBP-induced male reproductive effects were evaluated according to the lifestage of exposure

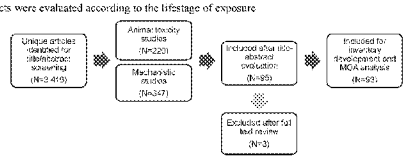


Figure 1. Abbreviated literature flow diagram

Disclaimer: The views expressed in this poster are those of the author(s) and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency
Office of Research and Development

Figure 2: Pathway for DBP-induced male reproductive effects after gestational exposure during MPW

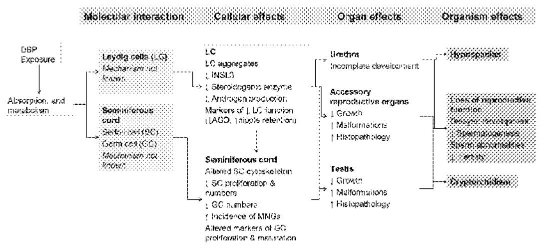
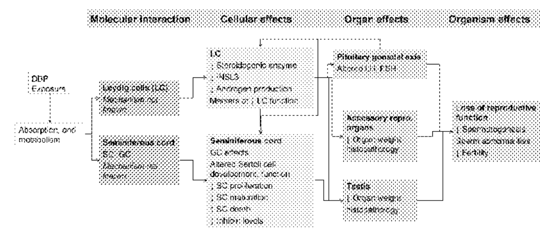


Figure 3: Pathway for DBP-induced male reproductive effects in post-natal lifestages



Results and discussion

Gestational exposure studies:

- Fetal rats appear more sensitive to DBP-induced anti-androgenic effects than are mice and may be more sensitive than other rodent species, non-human primates, and human fetal testis xenografts and ex-vivo tissue cultures.
- DBP-induced androgen-independent effects in the seminiferous cord (SC & GC) are conserved among most mammalian models (rats, rabbits and mice) and human xenografts.

Post-natal lifestage studies using peri-pubertal or sexually mature animals:

- DBP-induced Leydig cell effects are conserved in different mammalian species: (rats, rabbits, mice, gerbils, and guinea pigs, non-human primates [in-vivo and xenografts]).
- DBP-induced effects in the seminiferous cord (SC & GC) are also conserved among most mammalian models (rats, mice, and non-human primate [xenograft]).

Table 1: Preliminary cross-species coherence analysis for gestational effects

Key event	Animal in-vivo evidence	Animal (ex-vivo, xenograft)	Humans evidence (ex-vivo, xenograft)
Leydig cells (LCs)	No evidence		Not identified in studies
Sertoli cells (SCs) germ cells (GCs)	No evidence		Not identified in studies
LCs	■ Rat [ms] & rabbits [1] ■ Mice [1] □ Marmosets [1] & mice [3]	■ Rat xenograft [2] □ Rat ex-vivo* [2] □ Mice xenograft [1] □ Mice ex-vivo [1]	□ Human xenografts [3] □ Human ex-vivo [2]
SCs, GCs	■ SC and GC effects in rats [ms] ■ SC and/or GC effects in rabbits [1] mice [3] □ Marmoset [1]	■ Mice ex-vivo [2] ■ Mice xenograft [1]	■ Human xenografts [3] ■ Human ex-vivo [1]
Urethra	■ Rats [ms]		Not evaluated
Accessory reproductive organs	■ Rats [ms] & rabbits [1] □ Marmoset [1]	■ Rat xenograft [1]	□ Human xenograft [2]
Testis	■ Rats [ms], rabbits [1] & mice [2] □ Marmoset [1] & mice [3]		No evaluated
Organism effects reproductive functions	■ Rats [ms] & rabbits [1] □ Marmoset [1]		Not evaluated

■ – Evidence of response to exposure
□ – Evidence of no response (or reduced sensitivity) to exposure
[ ] – number of studies identified in the literature
ms – Many studies
* Both rat and mouse ex-vivo studies report no effect on basal T production, but gonadotropin-stimulated T was inhibited by exposure.

Table 2. Preliminary cross-species coherence analysis for effects in early post-natal lifestages

Key Event	Animal evidence (in-vivo)	Animal evidence (cell culture, xenograft)	Human evidence (ex-vivo, xenograft)
Leydig cells (LCs)	No evidence		No studies available
Sertoli cells (SCs), Germ cells (GCs)	No evidence		No studies available
LCs	■ Rats [17], mice [3], rabbits [1], marmoset [1] □ Rats [1], mice [1]	■ Cell culture models (rat [3], mouse [7] & dog [1]) ■ Rhesus monkey xenografts [1]	No studies available
SCs, GCs	■ Rats [32], mice [5] □ Marmoset [1]	■ Cell culture (rats [9]), mice [3], rhesus monkey xenografts [1]	No studies available
Pituitary gonadal axis	■ Rats [6] □ Rabbits [1], mice [2], rats [1]		No studies available
Accessory reproductive organs	■ Rats [10], rabbits [1], gerbils [1], mice [2] □ Mice [3], rats [6]	■ Rhesus monkey xenografts [1]	No studies available
Testis	■ rats [34], rabbits [1], mice [5], & guinea pigs [1] □ Hamsters [1], rats [5] mice [3], marmosets [2]	■ Rhesus monkey xenografts [1]	No studies available
Reproductive functions	■ Rats [12], rabbits [1], mice [5], & guinea pigs [1] □ mouse [2], rats [1]		No studies available

■ - Evidence of response to exposure
□ - Evidence of no response (or reduced sensitivity) to exposure
[ ] – number of studies identified in the literature

Selected references

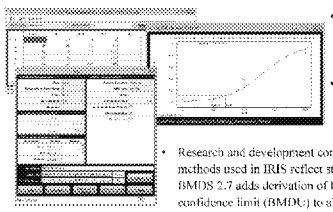
Albert O, and Jégou B. *Hum Reprod Update*. 2014. 20(2): 231-49.
Edwards et al., *Journal of Pharmacology and Experimental Therapeutics*. 2016, 356(1): 170-181.
Howdeshell KL, et al., *Environ Res*. 2008. 108(2): 168-76.
Makris SL, et al., *Birth Defects Res B Dev Reprod Toxicol*. 2008. 83(6): 530-46.
Boobis AR, et al., *Crit Rev Toxicol*. 2008. 38(2): 87-96.
Johnson KJ, et al., *Toxicol Sci*. 2012. 129(2): 235-48.
US EPA. (2009) An approach to using toxicogenomic data in US EPA human health risk assessments: a dibutyl phthalate case study.
US EPA. (2006) A framework for assessing health risk of environmental exposures to children.

EPA Dose-Response & Related Software - New & Future Developments

J. Allen Davis[1], Jeff Gift[2], David Farrar[3], Jay Zhao[1], and Matt Wheeler[3]

[1] US EPA, Office of Research and Development, National Center for Environmental Assessment - Cincinnati
[2] US EPA, Office of Research and Development, National Center for Environmental Assessment - Research Triangle Park
[3] Research Triangle Institute, Center for Chemical Safety and Health, Risk Science Group - Research Triangle Park

Benchmark Dose Software (BMDS 2.7 released 8/17)

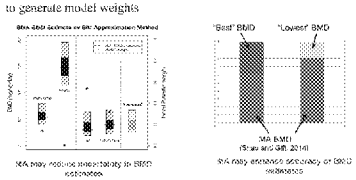


- Benchmark dose (BMD) method proposed by Crump (1984)
- Accepted as default dose-response modeling approach by US EPA (2012)
- Research and development continues to ensure methods used in IRIS reflect state-of-the-science, e.g., BMDS 2.7 adds derivation of BMD upper bound confidence limit (BMDL) to all models (USEPA 2017)

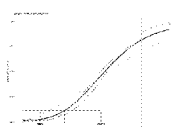
BMDS 3.0 - to be released in FY18

Bayesian Model Averaging

- EPA NCEA and NIOSH are developing Bayesian modeling averaging methods to address and/or account for model uncertainty
- Current methods for single model selection (i.e., AIC-based selection) have been shown to be inadequate (i.e., methods do not achieve nominal coverage rates)
- Current method uses maximum a posteriori estimation and Laplace approximations to generate model weights



- Method allows for assignment of model parameters and model weights, allowing for incorporation of biological or other prior information
- For example, information of a particular endpoint's mode of action may support weighting non-linear models more heavily than linear ones



$$\Pr(BMD \mid D) = \sum_{i=1}^{9} \pi_i \Pr(BMD \mid M_i, D)$$

Posterior Distribution of the BMD

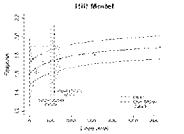
$$\alpha = \int_{-\infty}^{BMD_\alpha} \Pr(BMD \mid D) dBMD$$

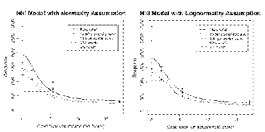
α^{th} percentile of the BMDL

U.S. Environmental Protection Agency
Office of Research and Development

BMDS 3.0 - to be released in FY18 (continued)

- **Hybrid Approach** – instead of using change in central tendency, the hybrid approach estimates a BMD using the percentage change in a population in the tail of the distribution
- Use of the hybrid approach for continuous data harmonizes benchmark responses between continuous and dichotomous data





- **Log-normality vs. Normality** – Shao and Gift (2013) determined that the distribution assumption has limited impact on the BMD estimates when the within dose-group variance is small
- BMDs defined using the hybrid approach are more sensitive to the distribution assumption

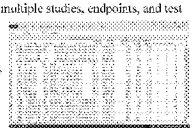
Categorical Regression (CatReg 3.1 released 6/17)

Categorical Regression

- Estimates the probability that a response occurs of a severity level, s, or greater given a concentration, C, and duration of exposure, T, as:

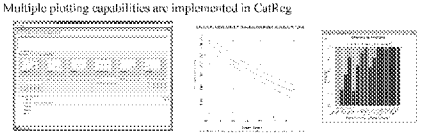
$$P(Y \geq s|C,T) = H[\alpha_s + \beta_{1S} * C + \beta_{2S} * T]$$

- CatReg allows for meta-analysis of data from multiple studies, endpoints, and test species (USEPA 2017; Milton et al., 2017)
- CatReg accounts for within study correlations (clustering) and allows for the stratification of model parameters to account for response differences across strata of data.



$$\Pr(Y \geq s|C,T,i) = H[\alpha_s + \gamma_i + \beta_1 j \times f_1(C) + \beta_2 k \times f_2(T)],$$
$$s = 1,2,\ldots,S,\ \ i = 1,2,\ldots,I,\ \ j = 1,2,\ldots,J,\ \ k = 1,2,\ldots,K$$

- CatReg incorporates hypothesis testing to allow users to determine the most appropriate form of the model (i.e, which variables should be stratified)
- Multiple plotting capabilities are implemented in CatReg

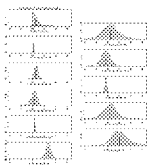


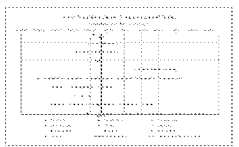
- U-shaped dose-response analysis could be added to future CatReg versions to facilitate assessment of toxicity from excess and deficiency (Milton et al., 2017)

Some Additional Related Developments and Plans

Probabilistic Meta-Analysis Methods for Meta-Analysis of Epidemiological Data

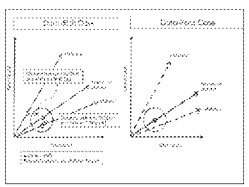
- Probabilistic meta-analysis dose-response methods have been proposed (NRC, 2008, 2013) to better assist risk management decision making
- Meta-analysis tools that allow for the combination of a multiple types of epidemiological studies using Bayesian statistics and hierarchical modeling have been developed to support future Agency health assessments





Mixture Similarity Tool (MIST)

- EPA Excel tool (MIST) based on Marshall et al. (2013)
- Data-Rich Case: Mixtures are similar when distance between reference and candidate mixture BMDs is less than radius of red circle
- Data-Poor Case: Simplifying assumptions to estimate distance via comparison of mixing proportions and weights for components of reference & candidate mixtures.



Addressing NRC Recommendations

New and future developments in dose-response modeling specifically address multiple recommendations provided by NRC (2014)

- "EPA should use formal methods for combining multiple studies and the derivation of IRIS toxicity values"
  - Both CatReg and meta-analysis tools for epidemiological data have been developed to increase IRIS' meta-analytical capabilities
- "Advanced analytic methods, such as Bayesian methods, for integrating data from dose-response assessments and deriving toxicity estimates are underused in the IRS program
  - Bayesian methods have recently been developed for use in IRIS assessments, including Bayesian model averaging and hierarchical Bayesian meta-regression approaches
- "Uncertainty analysis should be conducted systematically and coherently in IRIS assessments
  - Uncertainty analysis is supported by reporting entire confidence interval around BMD (BMDL – BMDU), which is done in the new model averaging method and CatReg

References

Marshall et al. (2013) An empirical approach to sufficient similarity: combining exposure data and in vitro toxicity data. *Risk Analysis* 33(9), pp. 1582-1595.

Milton et al. (2017) Modelling U-shaped dose-response curves for mutagenicity using categorical regression. *Mutagenicology*, 34, 217-225.

Shao, K. and Gift, J. S., 2013. Model uncertainty and Bayesian model averaged benchmark dose estimation for continuous data. *Risk Analysis*, 34(1), pp. 101-120.

US EPA, 2012, Benchmark Dose Technical Guidance Document, https://www.epa.gov/risk/benchmark-dose-technical-guidance

US EPA 2017a Benchmark Dose Software (BMDS) v. 2.7, https://www.epa.gov/bmds

US EPA 2017b Categorical Regression (CatReg) v. 3.1, https://www.epa.gov/bmds/catreg

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable based inks.

Evidence profile table for DIBP and male reproductive toxicity

Erin E. Yost[1], Xabier Arzuaga[2], Elizabeth Radke[2]
[1] US EPA National Center for Environmental Assessment, Research Triangle Park, NC; [2] US EPA National Center for Environmental Assessment, Washington, DC

The evidence profile table is a tool that complements the evidence integration narrative for human and animal data. Explanations for factors that increase or decrease confidence are provided in summaries.

Outcome*		Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings	Within stream evidence judgment	Inference across evidence streams	Overall conclusion
HUMAN STUDIES								
Testosterone (adult)		(All cross sectional studies) **Medium confidence** Meeker and Ferguson (2014) Pan et al., 2016 **Low confidence** Chang et al. (2015) Dan Hond et al. (2015)	• Consistency • Minimal risk of bias in medium confidence studies	• Few studies available	⊕⊕◯ MODERATE Inverse associations between DIBP exposure and testosterone levels in 3/4 studies (Meeker and Ferguson et al., 2014; Pan et al., 2016; Chang et al., 2015), 2 of which were statistically significant. No studies examined exposure-response gradient	⊕⊕◯ MODERATE Based on data for testosterone in adults, supported by slight evidence in other outcomes with low sensitivity and few available studies explaining lack of clear associations.	**Relevance of animal data to humans** •Role of androgen-dependent and –independent pathways in male reproductive system development, maturation, and function is conserved across mammalian species **Cross-stream coherence** •Testosterone is reduced with phthalate exposure in both humans and animals during different lifestages. **Susceptibility** •Developmental stages are particularly susceptible to perturbation by phthalates **Other relevant information** •Evidence from DBP, a structurally similar phthalate, provides robust evidence of male reproductive toxicity in humans, likely due to higher exposure levels and a larger number of studies	⊖⊕⊖ Overall conclusion that DIBP causes male reproductive toxicity, based on: 1) Robust evidence from oral exposure studies in rats and mice, with significant outcomes in gestational exposure studies at doses as low as 300 mg/kg-day. 2) Moderate evidence in human epidemiological studies of decreased testosterone in adult men with median metabolite concentrations in urine ranging from 7-48 ng/mL. Evidence for other outcomes was from populations with low urine metabolite concentrations, which reduced study sensitivity; and 3) Supporting mechanistic evidence demonstrating decreased testicular steroidogenesis and INSL-3. Evidence from animals is presumed relevant to humans. Lower level of evidence in humans can be explained by low sensitivity and few available studies.
Anogenital distance (AGD), semen parameters, pubertal development, time to pregnancy, hypospadias/cryptorchidism					⊖◯◯ SLIGHT			
ANIMAL STUDIES								
Gestational exposure	Testosterone	**High confidence** Borch et al. 2006 Furr et al. 2014 Hannas et al. 2011 Hannas et al. 2012 Howdeshell et al. 2008 Saillenfait et al. 2017 **Medium confidence** Wang et al., 2017	• Consistency • Exposure-response gradient • Effect size • Biological plausibility (support from mechanistic evidence) • Minimal risk of bias		⊕⊕⊕ ROBUST A dose-related decrease in testicular androgen levels or production (up to ~90% compared to control) was observed in all studies in rats and mice that evaluated this endpoint. Several of these studies also demonstrate decreased testicular expression of genes and proteins in the steroidogenesis pathway, which provides support for biological plausibility.	⊕⊕⊕ ROBUST Supported by consistency and coherence across outcomes, with mechanistic evidence (e.g. decreased testicular expression of **steroidogenic enzymes and INSL-3**) providing support for biological plausibility. The greatest weight of evidence came from gestational exposure studies, whereas postnatal exposure studies were limited by risk of bias concerns.		
	Male morphological development	**High confidence** Borch et al. 2006 Saillenfait et al. 2008 Saillenfait et al. 2008 Saillenfait et al. 2017 **Medium confidence** Wang et al., 2017	• Consistency within rat studies • Exposure-response gradient • Effect size • Biological plausibility • Minimal risk of bias		⊕⊕⊕ ROBUST All rat studies observed a dose-related increase in effects consistent with decreased testosterone and INSL-3, including increased time to puberty, decreased AGD, nipple retention, cryptorchidism, hypospadias, exposed os penis, and cleft prepuce. No effects on AGD were observed in mice (Wang et al. 2017).			
	Sperm evaluation and histopathological effects in testis or epididymis	**High confidence** Saillenfait et al. 2008 **Medium confidence** Borch et al. 2006 Wang et al., 2017	• Consistency • Exposure-response gradient • Effect size • Biological plausibility		⊕⊕⊕ ROBUST Adverse effects on the testis and/or sperm were observed in rats and mice, including a dose-related increased incidence of pathological lesions of the testis (Borch et al. 2006; Saillenfait et al., 2008), epididymal oligo- or azoospermia (Saillenfait et al. 2008), and decreased sperm concentration and motility (Wang et al. 2017).			
	Reproductive organ weight	**High confidence** Saillenfait et al. 2008 **Medium confidence** Wang et al., 2017	• Exposure-response gradient • Biological plausibility • Minimal risk of bias	• Few studies	⊕⊕◯ MODERATE Decreased reproductive organ weights were observed in rats (Saillenfait et al., 2008), whereas a consistent trend in testis weight was not observed in mice (Wang et al. 2017).			
Postnatal exposure	Testosterone				◯◯◯ INDETERMINATE			
	Sperm evaluation and histopathological effects in testis or epididymis	Low confidence Oishi and Hiraga 1980c Foster et al. 1981	• Consistency • Biological plausibility • Coherence with gestational exposure studies	• High risk of bias • Few studies	⊕⊖◯ MODERATE Rats were found to have increase testicular atrophy (Foster et al. 1981) and decreased spermatocytes and spermatogonia (Oishi and Hiraga 1980c).			
	Reproductive organ weight	**Medium confidence** Oishi and Hiraga 1980a Oishi and Hiraga 1980b Oishi and Hiraga 1980c Oishi and Hiraga 1980d **Low confidence** Foster et al. 1981 U. Rochester 1954 Zhu et al. 2010	• Consistency within rat studies • Biological plausibility • Coherence with gestational exposure studies	• High risk of bias in some studies • Unexplained inconsistency	⊕⊖◯ MODERATE In rats, a dose-related decrease in absolute testis weight was consistently observed (Oishi and Hiraga 1980c; d; Foster et al. 1981; University of Rochester 1954). In mice, Zhu et al. (2010) observed decreased testis weight in the highest dose group, whereas Oishi and Hiraga (1980a-b) observed increased testis weight.			

Comparison with findings from the National Academy of Sciences (NAS) systematic review of the low-dose toxicity of phthalates (2017)

Table 1: Summary of conclusions for DIBP from NAS 2017.

Endpoint	Initial hazard evaluation
Testosterone	Presumed human hazard • Based on high level of evidence from animal studies and inadequate evidence from human studies
AGD	Not classifiable • Based on inadequate evidence from human and animal studies
Hypospadias	Not classifiable • Based on inadequate evidence from human and animal studies

Both IRIS and NAS (2017) concluded that DIBP is likely to cause male reproductive toxicity in humans. However:

- NAS was only able to draw this conclusion for testosterone, based on the high level of evidence from rodent studies. Other endpoints (AGD and hypospadias) were determined to have inadequate evidence available.
- The IRIS systematic review was broader in scope (see Table 2) and was able to draw conclusions for a range of androgen-dependent and –independent male reproductive outcomes.

Table 2: Summary of major scoping differences between the IRIS and NAS systematic reviews of DIBP

	IRIS	NAS
Exposure	All life stages and dose levels	In utero exposure. Animal studies using a single dose ≥500 mg/kg-day are excluded.
Outcomes	Any male reproductive outcome	Testosterone, AGD, hypospadias

*Outcomes with slight or indeterminate evidence received a full systematic review, but were not significant contributors to the overall conclusion, so the details of the evidence are not provided here.

U.S. Environmental Protection Agency
Office of Research and Development

EPA United States Environmental Protection Agency

A New Bayesian Approach to Combining Different Species Data

Leonid Kopylev and Junyong Park
US EPA, Office of Research and Development, National Center for Environmental Assessment – Washington
University of Maryland at Baltimore County, Department of Mathematics and Statistics and ORISE Faculty Fellow at US EPA/ORD/NCEA Washington

Purpose and Scope

NRC (2014) recommended that IRIS develop the capacity to do Bayesian modeling of chemical hazards. In particular, NRC stated that "...more sophisticated Bayesian approaches have been proposed for combining dose-response estimates for multiple species and multiple chemicals (DuMouchel and Harris 1983; Jones et al. 2009). Those approaches might also be useful to EPA if guidance for selection of appropriate models and priors is developed."

In this research and development effort, EPA evaluated DuMouchel and Harris (1983) approach, developed alternative approach and applied it to the data in Jones et al. 2009

Background

DuMouchel and Harris (1983, JASA) were the first authors that addressed the problem of combining information for multiple species with a non-simplistic approach.

- Proposed a Bayesian approach to interspecies extrapolation
- Special attention to combining dose-response information
- Realized that they need subject matter expertise

DuMouchel and Harris (1983) realized that

- the species do not need to be restricted to humans and animals
- any type of data (including cell potency) is appropriate.
- a lot of toxicological experience is needed to figure out what chemicals dose-response information is combined.

Their ANOVA structure, however, assumes constant relative potency across species, which may not be the case in many examples.

A suggested model: Gaussian graphical model

Example (Jones et al. 2009; *Low birth weight*)

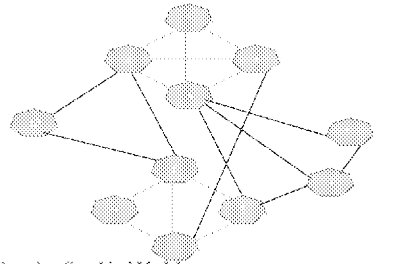
	Total THMs	Chloroform	BDCM	DBCM	Bromoform
Humans	1	2	3	4	12 (missing)
Rat S-D		5	6	7	8
Rat S-F			9		
Rabbit D-B		10			
Rabbit N-Z			11		

- Cell 1 - Cell 11 have the slope of regression model from log(dose) and log(response)
- Empty cells represent no-data

Assumptions

- We assume that species are related for the same chemical and different chemicals are related for the same species.
- We model dependence or relationship among different species and different chemicals through edges in Gaussian graphical model.
- We need to control dependence through prior probabilities for edges based on scientific knowledge rather than subjective choice.

Graphical representation of the Example



Gaussian Graphical Model

- Gaussian graphical model uses the inverse of covariance matrix called a precision matrix.
- Each component in a precision matrix represents the partial correlation between two nodes in the graphical model. Red edges shown for the same chemical data.
- No-edge between two nodes in graphical model is equivalent to partial correlation equal to 0.
- If there is no edge between cells i and j and the correlation is non-zero, then the nonzero correlation is due to all other data.

Formulation of the Bayesian graphical model

y: observed data

$$y \mid \beta, \Sigma \approx N(\beta, \Sigma)$$

Σ: Hyper Inverse Wishart Distribution

$e_{ij} \approx Bernoulli(p_{ij})$: edge between two nodes

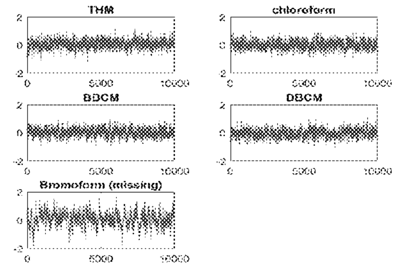
Prior probabilities on edges (representing existence of partial correlations)

- We give high prior probabilities to edges when two nodes have a close relationship.
- Such prior probabilities are still subjective, so they should be determined based on scientific knowledge to minimize subjectivity.

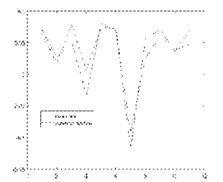
Results

95% Confidence Intervals of Human Data

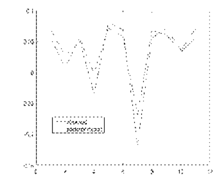
THM	Chloroform	BDCM	DBCM	Bromoform (missing)
0.0855	0.0234	0.0481	-0.0235	0.0878
(-.459,.652)	(-.536,.576)	(-.490,.578)	(-.571,.559)	(-.856,.998)



Validation of the Proposed method (Cross Validation)



Human data 1 - THM, 2 - Chloroform, 3 - BDCM (assumed to be missing), 4 - DBCM



Human data 1 THM, 2 Chloroform, 3 BDCM, 4–DBCM(assumed to be missing)

Comments on Results

- Estimate of missing value has more variation.
- When BDCM (or DBCM) is assumed to be missing, the posterior median of predicted values of human BDCM (or DBCM) is close to the observed value of BDCM (or DBCM).
- The patterns of the posterior medians are similar to those of the observed data.

Discussion and Future Directions

- We followed NRC (2014) recommendations on using Bayesian analysis and specifically investigated methodology proposed by DuMouchel and Harris (1983) and Jones et al. (2009).
- We proposed a new Bayesian method and validated recovery of missing human dose-response using Jones et al. (2009) data
- We will use simulation studies to validate our new method and consider its application to additional real data sets.
- We will consider extending the idea of graphical model to the area of combining DNA or RNA sequence data generated from different species.
- We will also consider application of the methodology to more data-poor examples that are more common in IRIS assessment work.

The views expressed in this poster are those of the authors and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency

U.S. Environmental Protection Agency
Office of Research and Development

Appendix E

Committee Findings Regarding 2014 Recommendations

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
1A	2	EPA needs to complete the changes in the IRIS process that are in response to the recommendations in the [2011] NRC formaldehyde report.	The 2014 report reviewed and encapsulated recommendations from the 2011 report, so the present committee focused its review on assessing progress made in implementing recommendations made by the 2014 report.	Workshop presentations, posters, and discussion Recent IRIS documents (such as plans, protocols, and assessments) and tools.
1B	2	[EPA needs to] specifically complete documents, such as the draft handbook, that provide detailed guidance for developing IRIS assessments. When those changes and the detailed guidance, such as the draft handbook, have been completed, there should be an independent and comprehensive review that evaluates how well EPA has implemented all the new guidance. The present committee is completing its report while those revisions are still in progress.	The revised handbook was not provided to the committee. EPA staff indicated that the handbook is under internal agency review and that its public release is expected in 2018. The agency further indicated that standard operating procedures might evolve as the IRIS program gains additional experience in performing systematic review and using emerging methods. The committee expects handbook revisions to be a continuing process, and EPA similarly characterizes the IRIS handbook as "evergreen." The committee observed that guidance for conducting newly planned IRIS assessments is contained in protocols, and EPA stated that some material currently in protocols might reside in the handbook. The amount of and need for overlap in the protocols and handbook could not be judged without seeing the handbook.	Slides 21–22 Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b]
2	2	EPA should provide a quality-management plan that includes clear methods for continuing assessments of the quality of the process. The roles of the various internal entities involved in the process, such as the CASTs, should be described. The assessments should be used to improve the overall process and the performance of EPA staff and contractors.	IRIS management has taken multiple steps to ensure high-quality management, including the creation of expertise-specific work groups, systematic-review work groups, and other intermediate structures to improve the quality of the IRIS assessments. EPA has also used the SAB Chemical Assessment Advisory Committee to review IRIS assessments. Funding for contractors has decreased.	Slides 7–10, 151 The GAO audit of the IRIS program indicates that improvements in program management have occurred (Slide 10)
3	2	When extracting data for evidentiary tables, EPA should use at least two reviewers to assess each study independently for risk of bias. The reliability of the independent coding should be calculated; if there is good agreement, multiple reviewers might not be necessary.	EPA uses two people to extract data and, when needed, involves a third person to resolve conflicts. EPA also uses two people to complete the risk-of-bias evaluation.	Slide 39 Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (p. 17, line 1; p. 30, lines 18–20)[b]
4A	2	EPA should continue its efforts to develop clear and transparent processes that allow external stakeholder input early in the IRIS process.	EPA has adopted the process of soliciting public comment early through the release of assessment plans and protocols for public comment.	IRIS Web site Slides 24–25, 29
4B	2	[EPA] should develop communication and outreach tools that are tailored to meet the needs of the various stakeholder groups. For example, EPA might enhance its engagement with the scientific community through interactions at professional-society meetings, advertised workshops, and seminars. In contrast, greater use of social media might help to improve communications with environmental advocacy groups and the public.	Although this recommendation was not discussed specifically with EPA, the agency has worked in the past with the National Academies to identify experts that could provide input at IRIS workshops. The IRIS Web site provides features for sharing information via social-media tweets and Facebook. The calendar feature clearly indicates the schedule for public engagement events on IRIS assessments. EPA staff also discussed data- and tool-sharing with stakeholders to increase understanding and accessibility of systematic-review practices used to develop IRIS assessments.	IRIS Web site Slide 15

(Continued)

Continued

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
5	2	Similar to other EPA technical-assistance programs, EPA should consider ways to provide technical assistance to under-resourced stakeholders to help them to develop and provide input to the IRIS program.	This recommendation was not discussed specifically with EPA.	
6	2	The stopping rules should be explicit and transparent, should describe when and why the window for evidence inclusion should be expanded, and should be sufficiently flexible to accommodate truly pivotal studies. Such rules could be included in the preamble.	The issue of stopping rules was not specifically discussed, but the IRIS program has completed a rapid review of chloroprene, and this is consistent with this recommendation.	Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c]
7	2	Regarding promotion of efficiencies, EPA should continue to expand its efforts to develop computer systems that facilitate storage and annotation of information relevant to the IRIS mission and to develop automated literature and screening procedures, sometimes referred to as text-mining.	EPA has made considerable progress in developing and upgrading the Health and Environmental Research Online (HERO) database and the Health Assessment Workspace Collaborative (HAWC) computer system to facilitate storage and annotation of data. Those systems are not subject to third party control. EPA is also using other software systems, including the Sciome Workbench for Interactive computer-Facilitated Text-mining (SWIFT) and related products for text-mining.	Workshop Demonstrations Slides 36, 92–116
8	2	More details need to be provided on the recognition and applications of expert judgment throughout the assessment-development process, especially in the later stages of the process. The points at which expert judgment is applied should be identified, those applying the judgment should be listed, and consideration should be given to harmonizing the use of expert judgment at various points in the process.	EPA has developed guided expert judgment to synthesize evidence on the basis of modified Bradford Hill criteria and for integrating evidence across data streams. The agency has developed working groups with expertise (such as PBPK) that can be applied to the assessment process. The draft chloroform protocol identified some situations when expert judgment will be used, including evaluation of studies to identify characteristics that indicate how informative the results are (p. 16, line 21) to perform outcome-specific study evaluations (p. 16, line 24).	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b] Slides 8, 48, 69–86
9	3	EPA should establish a transparent process for initially identifying all putative adverse outcomes through a broad search of the literature. The agency should then develop a process that uses guided expert judgment to identify the specific adverse outcomes to be investigated, each of which would then be subjected to systematic review of human, animal, and in vitro or mechanistic data.	EPA has developed assessment plans that provide information about the scoping and problem formulation process. The plans are developed by using expert judgment and input from EPA regional offices and other stakeholders. Each assessment plan identifies the specific aims of the systematic review and the PECO statement.	IRIS Assessment Plan for Chloroform (Scoping and Problem Formulation Materials)[d]
10	3	For all literature searches, EPA should consult with an information specialist who is trained in conducting systematic reviews.	EPA staff indicated that they use an information specialist.	EPA protocol provides the name of the HERO librarian (see chloroform protocol, page vii);[b] that person has an MS in library and information science
11	3	EPA should include protocols for all systematic reviews conducted for a specific IRIS assessment as appendixes to the assessment.	The IRIS program has developed draft systematic-review protocols that are undergoing public comment before being made final. The protocols contain many of the elements identified by the 2014	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation), Appendix A[b]

			report as meeting best practices defined by IOM. Furthermore, the chloroprene reassessment included as appendixes the literature-search strategy and approaches for evaluating risk of bias in epidemiology and other human studies. The study objective, PECO statement, and methods used to search and screen the literature and evaluate studies were included in the main body of the report. That approach is consistent with this recommendation. The committee expects that some items found in the protocol can be addressed in the handbook. Including the analysis plan in the systematic-review protocols might lead to additional amendments to the protocol that could be minimized if they used a separate analysis plan.	Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c]
12	4	The trajectory of change needs to be maintained.	The IRIS program has been responsive to the recommendations made in the 2014 report and is continuing the trajectory of change. The changes appear to have accelerated with the recruitment of new NCEA and IRIS leadership.	Workshop presentations, posters, and discussion Recent IRIS documents (such as plans, protocols, and assessments) and tools
13	4	The current process can be enhanced with more explicit documentation of methods. Protocols for IRIS assessments should include a section on evidence identification that is written in collaboration with information specialists trained in systematic reviews and that includes a search strategy for each systematic-review question being addressed in the assessment. Specifically, the protocols should provide a line-by-line description of the search strategy, the date of the search, and publication dates searched and, as noted in Chapter 3, explicitly state the inclusion and exclusion criteria for studies.	EPA systematic-review protocols contain descriptions of how evidence will be identified, including relevant search terms and databases to be queried. The protocols also include descriptions of inclusion and exclusion criteria.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation), Table 2 (p. 9), p. 12, Appendix A[b] Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c]
14	4	Evidence identification should involve a predetermined search of key sources, follow a search strategy based on empirical research, and be reported in a standardized way that allows replication by others. The search strategies and sources should be modified as needed on the basis of new evidence on best practices. Contractors who perform the evidence identification for the systematic review should adhere to the same standards and provide evidence of experience and expertise in the field.	EPA systematic-review protocols contain descriptions of how evidence will be identified, including relevant search terms and databases to be queried.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation), Appendix A[b]
15	4	EPA should consider developing specific resources, such as registries, that could be used to identify and retrieve information about toxicology studies reported outside the literature accessible by electronic searching. In the medical field, clinical-trial registries and US legislation that has required studies to register in ClinicalTrials.gov have been an important step in ensuring that the total number of studies that are undertaken is known.	This recommendation goes beyond the scope of the IRIS program and therefore was not addressed by the committee during its review. Systematic-review protocols indicate that IRIS assessments include only publicly accessible, peer-reviewed information, which should be available through the databases identified by the IRIS program.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (p. 11, line 14)[b]

(Continued)

Continued

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
16	4	EPA is encouraged to use at least two reviewers who work independently to screen and select studies, pending an evaluation of validity and reliability that might indicate that multiple reviewers are not warranted. It is important that the reviewers use standardized procedures and forms.	EPA uses two persons to screen studies. Screeners use a structured form based on the PECO in DistillerSR.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (p. 12, lines 11–16)[b] Slide 39
17	4	EPA should engage information specialists trained in systematic reviews in the process of evidence identification, for example, by having an information specialist peer review the proposed evidence-identification strategy in the protocol for the systematic review.	The IRIS assessment team includes an information specialist. The specific tasks completed by that person are not clear. It is hoped that the handbook will clearly define the roles that the person has in the IRIS process.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b]
18	4	EPA should encourage and support research on reporting biases and other methodologic topics relevant to the systematic-review process in toxicology.	EPA is supporting and encouraging research through its collaborative efforts described at the workshop. The committee expects EPA research in this field to emerge as the IRIS program continues to develop expertise in systematic-review method development.	Slides 79, 91, 149, 145, 150
19	4	EPA should continue to document and standardize its evidence-identification process by adopting (or adapting, where appropriate) the relevant IOM standards described in Table 4-1. It is anticipated that its efforts will further strengthen the overall consistency, reliability, and transparency of the evidence-identification process.	Appropriate tools and methods for evidence identification were described and are being used.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b] Workshop presentations
20A	5	To advance the development of tools for assessing risk of bias in different types of studies (human, animal, and mechanistic) used in IRIS assessments, EPA should explicitly identify factors, in addition to those discussed in this chapter, that can lead to bias in animal studies—such as control for litter effects, dosing, and methods for exposure assessment—so that these factors are consistently evaluated for experimental studies.	The draft chloroform protocol describes the domains to be considered in the evaluation of epidemiology studies and animal toxicity studies. Domain ratings and their descriptions have also been provided. EPA also presented heat maps of risk-of-bias analyses for studies performed by the IRIS program.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (Tables 5–6)[b] Slides 53, 55
20B	5	Likewise, EPA should consider a tool for assessing risk of bias in in vitro studies.	The 2014 report noted that few tools were available for assessing risk of bias in in vitro studies. Fully developed tools that meet the needs of the IRIS program are not available. EPA is exploring adaptations of existing tools for its purpose.	Slide 78
21A	5	When considering any method for evaluating individual studies, EPA should select a method that is transparent, reproducible, and scientifically defensible.	EPA has adopted systematic-review methods that are transparent and scientifically defensible.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b] Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c] Slides 50–63

21B	5	Whenever possible, there should be empirical evidence that the methodologic characteristics that are being assessed in the IRIS protocol have systematic effects on the direction or magnitude of the outcome. The methodologic characteristics that are known to be associated with a risk of bias should be included in the assessment tool. Additional quality-assessment items relevant to a particular systematic-review question could also be included in the EPA assessment tool.	EPA is using and adapting risk-of-bias tools appropriately.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b] Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c] Slides 52–63 Posters D-4, D-5, D-9
22	5	EPA should carry out, support, or encourage research on the development and evaluation of empirically based instruments for assessing bias in human, animal, and mechanistic studies relevant to chemical-hazard identification. Specifically, there is a need to test existing animal-research assessment tools on other animal models of chemical exposures to ensure their relevance and generalizability to chemical-hazard identification. Furthermore, EPA might consider pooling data collected for IRIS assessment to determine whether, among various contexts, candidate risk-of-bias items are associated with overestimates or underestimates of effect.	EPA is supporting and encouraging research through its collaborative efforts described in the workshop.	Slides 145, 149
23	5	Although additional methodologic work might be needed to establish empirically supported criteria for animal or mechanistic studies, an IRIS assessment needs to include a transparent evaluation of the risk of bias of studies used by EPA as a primary source of data for the hazard assessment. EPA should specify the empirically based criteria it will use to assess risk of bias for each type of study design in each type of data stream.	EPA has adapted existing risk-of-bias tools for its use. Draft protocols describe the domains to be considered in the evaluation of epidemiology studies and animal toxicity studies. Domain ratings and their descriptions have also been provided. EPA also presented heat maps of risk-of-bias analyses for studies performed by the IRIS program. Tools have not been developed for mechanistic studies.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (Tables 5 and 6)[b] Slides 53, 78
24	5	To maintain transparency, EPA should publish its risk-of-bias assessments as part of its IRIS assessments. It could add tables that describe the assessment of each risk-of-bias criterion for each study and provide a summary of the extent of the risk of bias in the descriptions of each study in the evidence tables.	EPA presented example heat maps of risk-of-bias analyses for studies performed by the IRIS program. The heat maps have been included in a recent assessment.	Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment (Figure 2)[c]
25	5	EPA should develop terminology for potential sources of bias with definitions that can be applied during systematic reviews.	EPA has adapted existing risk-of-bias tools for its use. The draft chloroform protocol describes the domains to be considered in the evaluation of epidemiology studies and animal toxicity studies. Reporting bias was not included as a domain for epidemiology studies, and its omission is not consistent with standard systematic-review methods for assessing risk of bias.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (Tables 5–6)[b] Slides 55, 57

(Continued)

Continued

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
26	5	Funding sources should be considered in the risk-of-bias assessment conducted for systematic reviews that are part of an IRIS assessments	EPA documents funding source, but it is unclear how the data are used.	Workshop discussion
27A	5	EPA should contact investigators to obtain missing information that is needed for the evaluation of risk of bias and other quality characteristics of included studies.	Investigators are contacted on a case-by-case basis that depends partly on the expected effect of the missing data. IRIS systematic-review protocols also indicate that decisions are made on an assessment-specific basis. If the information is not reported, it is generally not useful to reach out to the study authors. However, if missing study details could change confidence in study conclusions, efforts should be made to contact the study authors. Outreach to study authors is documented and considered unsuccessful if researchers do not respond to multiple e-mail or phone requests within a reasonable period.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (Table 6 (p. 25); p. 18, line 41)[b]
27B	5	The committee expects that, as happened in the clinical literature in which additional reporting standards for journals were implemented (Turner et al. 2012), the reporting of toxicologic research will eventually improve as risk-of-bias assessments are incorporated into the IRIS program. However, a coordinated approach by government agencies, researchers, publishers, and professional societies will be needed to improve the completeness and accuracy of reporting toxicology studies in the near future.	This recommendation goes beyond the scope of the IRIS program and therefore was not addressed during this review.	
28	5	The risk-of-bias assessment of individual studies should be carried forward and incorporated into the evaluation of evidence among data streams.	The results of the evaluation of individual studies are a critical component of the current evidence synthesis processes and integration frameworks. Risk of bias is one factor that EPA uses to determine an overall study confidence rating for epidemiology and animal toxicity studies. High- or medium-confidence studies are favored for quantitative dose–response analysis.	Slides 66, 54, 71–73, 81
29	6	EPA should continue to improve its evidence-integration process incrementally and enhance the transparency of its process. It should either maintain its current guided-expert-judgment process but make its application more transparent or adopt a structured (or GRADE-like) process for evaluating evidence and rating recommendations along the lines that NTP has taken. If EPA does move to a structured evidence-integration process, it should combine resources with NTP to leverage the intellectual resources and scientific experience in both organizations. The committee does not offer a preference but suggests that EPA consider which approach best fits its plans for the IRIS process.	The IRIS process continues to use a guided expert judgment process, but structured sets of categorical criteria for decision-making within that process are more explicitly defined.	Slides 67, 79–86

30	6	EPA should expand its ability to perform quantitative modeling of evidence integration; in particular, it should develop the capacity to do Bayesian modeling of chemical hazards. That technique could be helpful in modeling assumptions about the relevance of a variety of animal models to each other and to humans, in incorporating mechanistic knowledge to model the relevance of animal models to humans and the relevance of human data for similar but distinct chemicals, and in providing a general framework within which to update scientific knowledge rationally as new data become available. The committee emphasizes that the capacity for quantitative modeling should be developed in parallel with improvements in existing IRIS evidence-integration procedures and that IRIS assessments should not be delayed while this capacity is being developed.	EPA illustrated its use of meta-analysis of human and animal studies for evidence integration. Bayesian methods are being explored to help to characterize uncertainty and to combine evidence to identify hazard. New methods and assays are increasingly being evaluated quantitatively.	Slide 130 Posters provided examples that show how EPA uses new approach methods as part of a chemical assessment process
31	6	EPA should develop templates for structured narrative justifications of the evidence-integration process and conclusion. The premises and structure of the argument for or against a chemical's posing a hazard should be made as explicit as possible, should be connected explicitly to evidence tables produced in previous stages of the IRIS process, and should consider all lines of evidence (human, animal, and mechanistic) used to reach major conclusions.	The 2017 Toxicological Profile for Benzo[a]pyrene shows well-developed evidence tables that support the structured narrative and conclusion regarding carcinogenicity. For other effects, the evidence is described as ranging from "strongest evidence for human hazards" to "less robust evidence." Workshop discussion and the chloroform protocol show progress in template development. EPA staff stated that the approach to standardization of hazard descriptors for noncancer effects is being tested and discussed in the agency.	Slides 80–86 2017 IRIS Toxicological Profile for Benzo[a]pyrene[e] Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b]
32	6	Guidelines for evidence integration for cancer and noncancer end points should be more uniform.	Although EPA has not developed these guidelines, the issue goes beyond the IRIS program with respect to agency procedures. However, the IRIS program has developed frameworks for evidence integration and is testing and discussing how conclusions should be summarized.	
33	7	EPA should develop criteria for determining when evidence is sufficient to derive toxicity values. One approach would be to restrict formal dose-response assessments to when a standard descriptor characterizes the level of confidence as medium or high (as in the case of noncancer end points) or as "carcinogenic to humans" or "likely to be carcinogenic to humans" for carcinogenic compounds. Another approach, if EPA adopts probabilistic hazard classification, is to conduct formal dose-response assessments only when the posterior probability that a human hazard exists exceeds a predetermined threshold, such as 50% (more likely than not likely that the hazard exists).	Progress has been made. Quantitative toxicity values are restricted to studies with strongest conclusions for a human health effect (for cancer, a descriptor of *Known*) or a moderately strong conclusion for a human health effect (for cancer, a descriptor of *Likely*). Criteria are not provided for inclusion of studies that are considered on a case-by-case basis when a weaker conclusion regarding a human health effect (for cancer, a descriptor of *Suggestive*) is reached. IRIS has not produced final descriptors for noncancer effects and mechanistic studies other than review and application of PK/PBPK models.	Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c] 2017 IRIS Toxicological Profile for Benzo[a]pyrene[e] Slides 131–133

(Continued)

Continued

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
34	7	EPA should continue its shift toward the use of multiple studies rather than single studies for dose-response assessment but with increased attention to risk of bias, study quality and relevance in assessing human dose-response relationships. For that purpose, EPA will need to develop a clear set of criteria for judging the relative merits of individual mechanistic, animal, and epidemiologic studies for estimating human dose-response relationships.	Progress has been made toward using multiple studies or end points and comparing multiple candidate toxicity values. IRIS assessments provide one or more candidate toxicity values for use by risk managers. The IRIS program considers the quality of studies when deciding which studies will be advanced for quantitative dose–response modeling; studies rated as having medium or high confidence will be advanced for dose–response considerations. Other study attributes—such as relevance of a species to humans, relevance of an exposure route, and susceptibility—might also be considered. EPA is developing new tools for making and visualizing comparisons. EPA recognizes that there is no one-size-fits-all sets of criteria for inclusion of mechanistic studies, but the criteria for evaluating PK/PBPK models and how they are applied in dose–response and toxicity-value determinations are a good start.	Slides 62, 130–135, 142–146 2012 IRIS Toxicological Review of Tetrachloroethylene[d] Workshop demonstrations of HAWC and SWIFT
35	7	EPA should use formal methods for combining multiple studies and the derivation of IRIS toxicity values with an emphasis on a transparent and replicable process.	IRIS has begun to develop and apply tools in response to this recommendation. EPA presented two demonstrations for meta-regression and Bayesian approaches that showcase the agency efforts. EPA has not presented criteria for when and how new tools should be used. Tool development and application will be a continuing process that requires sustained resources and continued capacity-building.	Slide 140 Case studies provided for alternative dose estimates (posters D-2, D-10)
36	7	EPA should clearly present two dose-response estimates: a central estimate (such as a maximum likelihood estimate or a posterior mean) and a lower-bound estimate for a POD from which a toxicity value is derived. The lower bound becomes an upper bound for a cancer slope factor but remains a lower bound for a reference value.	EPA indicated that this approach is now standard procedure. Several examples were presented that show comparisons between BMDs and BMDLs and demonstrate how key studies compare with other supporting studies	Slides 134, 135; posters
37	7	As the IRIS program evolves, EPA should develop and expand its use of Bayesian or other formal quantitative methods in data integration for dose-response assessment and derivation of toxicity values.	Demos show the beginning stage of IRIS efforts on applications of Bayesian methods. EPA has not yet developed criteria for when and how new tools should be used. New research is under way to address New Approach Methods, such as data-mining, cheminformatics, high-throughput exposure modeling and toxicokinetics, and visualization tools.	Case studies (Poster D-10) Slides 136, 139, 140, 143–146

38	7	Uncertainty analysis should be conducted systematically and coherently in IRIS assessments. To that end, EPA should develop IRIS-specific guidelines to frame uncertainty analysis and uncertainty communication. Moreover, uncertainty analysis should become an integral component of the IRIS process.	Efforts are beginning with development of model-averaging approaches and adoption of WHO/IPCS guidance for reporting toxicity values and their uncertainty. IRIS specific guidance has yet to be developed because tools and approaches remain under development.	Cooper et al. (2016)[g] Slides 137, 138 Case studies (Poster D-9)

[a]NRC. 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press.

[b]EPA. 2018. Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) [CASRN 67-66-3]. EPA/635/R-17/486. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

[c]Orme-Zavaleta, J. 2018. Response to the Request for Correction (RFC). Letter to Robert Holden, Liskow & Lewis, New Orleans, LA, from Jennifer Orme-Zavaleta, Principal Deputy Assistant Administrator for Science, Office of Research and Development, Washington, DC, January 25, 2018; Attachment 1. EPA Response to the Denka Performance Elastomers (DPE) Request for Correction of the Toxicological Review of Chloroprene (CAS No. 126-99-8) In Support of Summary Information on the Integrated Risk Information System (IRIS); Attachment 2. Systematic Review of Chloroprene [CASRN 126-99-80] Studies Published Since 2010 IRIS Assessment to Support Consideration of the Denka Request for Correction (RFC). January 2018 [online]. Available: https://www.epa.gov/sites/production/files/2018-01/documents/epa_repsonse_to_mr._holdren_jan_25_2018_complete.pdf [accessed February 9, 2018].

[d]EPA. 2017a. IRIS Assessment Plan for Chloroform [CASRN 67-66-3]. EPA/635/R-17/330. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

[e]EPA. 2017b. Toxicological Review of Benzo[a]pyrene. EPA/635/R-17/003Fa. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0136tr.pdf [accessed February 14, 2018].

[f]EPA. 2012. Toxicological Review of Tetrachloroethylene (Perchloroethylene) [CAS No. 127-18-4]. EPA/635/R-80/011F. U.S. Environmental Protection Agency, Washington, DC. February 2012 [online]. Available: https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0106tr.pdf [accessed February 27, 2018].

[g]Cooper, G.S., R.M. Lunn, M. Ågerstrand, B.S. Glenn, A.D. Kraft, A.M. Luke, and J.M. Ratcliffe. 2016. Study sensitivity: Evaluating the ability to detect effects in systematic reviews of chemical exposures. Environ. Int. 92-93:605-610.